SVENSKA GEOFYSISKA FÖRENINGEN

VOLUME 7, NUMBER 1 **Tellus** FEBRUARY 1955

A QUARTERLY JOURNAL OF GEOPHYSICS

A Theoretical and Experimental Study of Cellular Convection in Rotating Fluids

YOSHINARI NAKAGAWA and PAUL FRENZEN, Hydrodynamics Laboratory, University of Chicago

(Manuscript received, September 26, 1954)

Abstract

The stabilizing and cell-shape changing influence of a basic rotation upon thermally produced cellular circulations are developed in a first order, perturbation theory. The theoretical possibility of a special "overstable" oscillating mode in certain fluids under the influence of rotation is recognized. In this preliminary report, the theory obtained for ordinary convection is compared to results obtained in laboratory experiments with rotating cylinders of water, and the existence of the overstable oscillating mode is demonstrated experimentally. Several possible applications of the theoretical results to atmospheric convection are developed by means of numerical examples, including models that are consistent with the observed association of deep convection with cyclonic relative vorticity and the distribution of clouds in regularly spaced clear and cloudy areas.

Beginning with what was probably the first reported observation of the phenomenon by Weber, who in 1855 described cells with descending currents in their centers which he had seen in a thin layer of alcohol and water under a microscope, cellular convective circulations have been the subject of many investigations from the viewpoints of both experiment and theory. In 1882, James Thomson made the famous observation of "tesselated structure" in a tub of warm, soapy water standing in the yard of an inn. Studies under the controlled conditions of the laboratory began with the classical papers of HENRI BENARD (1900, 1901), the term "Benard cells" often being used in appreciation of his pioneer work. These early investigations plus several of the more recent studies have been summarized by BRUNT (1937, 1951).

LORD RAYLEIGH (1916) contributed the basic theoretical treatment. He obtained an expression determining the degree of instability necessary to overcome the dissipative influences of viscosity and heat conduction in terms of a non-dimensional parameter now known as the Rayleigh number (equation (21) below). As STOMMEL (1947) has pointed out, all subsequent analyses have been essentially refinements and extensions of Rayleigh's original work.

A few writers have considered the theoretical influence of rotation upon cellular convection. JEFFREYS (1928), one of the first, pointed out that "the effect of rotation is to maintain stability" and noted that high viscosities could actually aid instability by preventing the dominance of the stabilizing effect of rotation. Stommel refers to unpublished work in which he apparently obtained expressions similar to

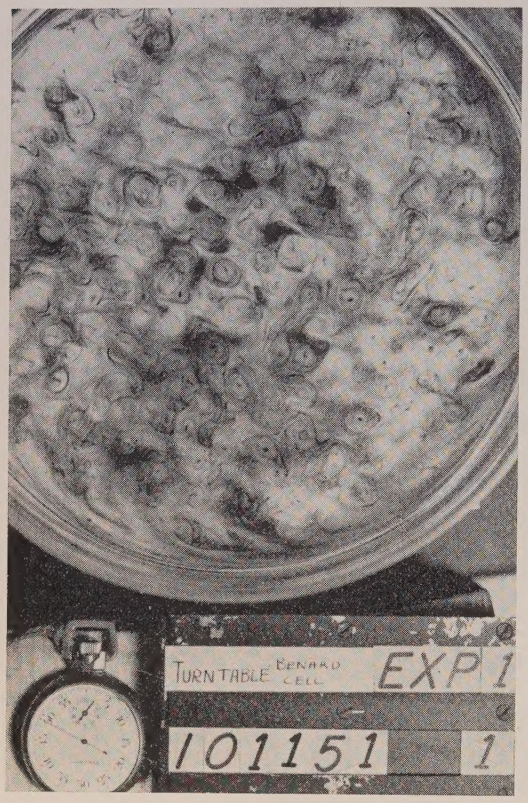


Fig. 1. Top view of convection cells in water driven by evaporation cooling from above. Water 24.3° C; air 22.5° C; depth 2.7 cm; rotating 10.2 rpm; Taylor number 2.4×10^6. The tracer is nigrosine ink introduced into the top layers. (Approx. one third size.)

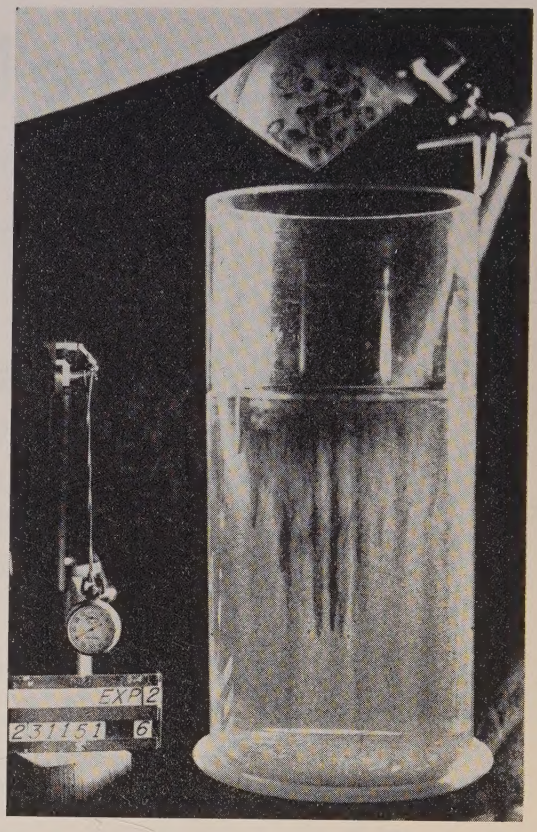


Fig. 2. Side view of convection cells in water driven by evaporation cooling from above. Water 25.7° C; air 22.7° C; depth 31 cm; rotating 25 rpm; Taylor number 2.5×10^{11}. Ink descending from the upper layers in the cell cores.

equations (13) and (15) below, for he mentions "an eight order partial differential equation in w and a cubic in n". He also concluded that rotation inhibits convection and added that the effects are negligible if the depth of the fluid layer is less than $(\nu/\Omega)^{1/2}$, where ν is the kinematic viscosity and Ω is the angular velocity.

CHANDRASEKHAR (1953) has covered the theoretical problem of the influence of rotation upon convection quite thoroughly. His recent work, containing in part the solution included in this paper, came to the authors' attention after the present analysis had been completed by Nakagawa. Chandrasekhar gives explicit expressions for the critical Rayleigh numbers necessary for convection in the case of two free bounding surfaces and approximate critical Rayleigh numbers for the case of two rigid and for the case of one free and one rigid boundary.

In the past, considerable speculation in cellular convection studies has been concerned with the observed directions of the vertical circulations. Early observations seemed to indicate that cells in liquids formed with ascending cores (with the exception of Weber, above), whereas cells in gases universally exhibited descending central cores. This last considerably damaged attempts to incorporate Benard cells in meteorological convection theory. Explanations based on the opposite variations of viscosity with temperature in liquids and gases were advanced for this apparent difference between convection cells in the two media.

Two investigators, Chandra and Dassanayake (summarized by BRUNT, 1937, 1951) separately

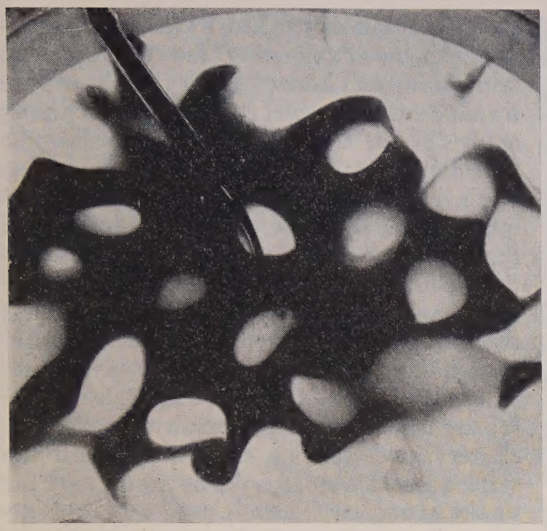


Fig. 3. Top view of holes punched in a pool of ink on the bottom of a cylinder of water by descending cores of cells driven by evaporation cooling. Water 24.3° C; air 23.5° C; depth 19 cm; rotating 10.0 rpm; Taylor number 5.7×10^9. (Approx. one third size.)

Fig. 4. Side view of ink rising from the bottom of a cylinder of water in the ascending cores of cells driven by heating from below. Water 23.5° C; air 24.0° C; depth 18 cm; rotating 10.9 rpm; Taylor number 5.5×10^9; heating plate input 5 watts. (Approx. full size.)

demonstrated a special form of convection in gases in which ascending central currents did occur; but this "columnar type" was never observed in gaseous layers deeper than eight millimeters. Further, measurements of the vertical temperature gradient in these shallow layers indicated that the circulation was fundamentally different from the usual Benard-Rayleigh type of convection. A much smaller vertical temperature gradient was required to initiate the ascending columnar circulations than that called for by the Rayleigh criterion. SUTTON (1950) has suggested that this special mode is the result of instability in the shallow boundary layer above the heated surface and that it becomes the principal mode when the overall fluid depth is of the order of the boundary layer depth.

Convection cells in rotating laboratory models were first observed at almost the same time by two individuals working independently at widely separated locations, by D. Fultz at the University of Chicago and by Y. Nakagawa at the University of Tokyo. In the Chicago experiments, fields of cells with descending central cores were seen to occur in rotating cylinders of water whose temperature was very close to that of the room (Figs 1 and 2). Evaporation cooling at the top surface provided the small degree of thermal instability necessary to drive the cells. They did not form in water more than a degree or two cooler than the room, while motions in water several degrees warmer than the air resembled the long wave and jet patterns observed in the "dishpan" experiments (CORN and FULTZ, 1954; FULTZ, 1952, 1954), because of horizontal temperature gradients set up by cooling at the walls of the container. In Tokyo, Nakagawa found cells with ascending central cores in rotating cylinders of water gently heated from below.

In every experiment run during the series reported in detail later, convection cells in rotating cylinders of water driven by cooling from above, either by evaporation or by contact with an external heat sink, exhibited

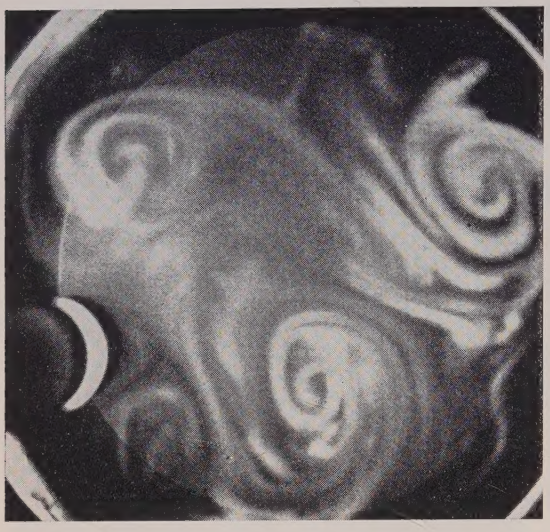


Fig. 5. Top view of an illuminated cross section 15 cm from the bottom of a rotating cylinder of air. The smoke tracer is converging and rotating cyclonically in the lower portions of cells with ascending cores driven by heating from below. Temperature 27° C; depth 51 cm; rotating 5.3 rpm; Taylor number 2.9×10^8; heating plate input 1 watt. (Approx. one third size.)

descending central cores, while all of those driven by heating from below had ascending cores. The direction of the horizontal circulation at the top of each cell was that appropriate to the conservation of angular momentum in the divergent or convergent field, i.e., cyclonic for cells with descending cores and anticyclonic for those with ascending cores. The difference is strikingly illustrated in Figs 3 and 4. Here, the conditions of the two experiments are almost identical; the Taylor number (see eqn. (22)) for both experiments is approximately 5.6×10^9. The slightly shallower depth in Fig. 4 (necessitated by the introduction of an aluminium plate bottom, in order to lift the base of the cells clear of the heating panel frame for a profile view) was compensated by an increase in the rate of rotation. In Fig. 3 evaporative cooling at the top surface has produced cells with descending cores that have punched holes in a pool of ink at the bottom, while in Fig. 4, a similar pool of ink feeds the ascending cores of cells driven by heating from below.

A series of unique observations of Benard-Rayleigh cells with ascending cores in air were made in a fifty-one centimeter deep, rotating cylinder heated from below. Cross sections of the circulation at selected levels were photographed by illuminating the tracer smoke with an intense, horizontal slit of light. At the fifteen centimeter level (Fig. 5) cyclonic spirals are visible in a convergent field feeding the ascending cores. By the time the smoke has reached the upper levels of the cylinder it is quite diffuse and only the general outlines of the cells can be seen (Fig. 6). But visual observations determined that here the smoke in the large vortices was rotating anticyclonically; therefore the field of motion was divergent. Descending cores were observed in air cells when the rotating cylinder was cooled at the top by contact with a water bath. Relatively clearer air descending from the cooled top surface punched small holes down through the smoke at regular intervals. In a few cases the descending cores could be seen extending almost to the bottom of the fifty centimeter cylinder, but they were too indistinct to be photographed.

The results of these qualitative experiments can be summarized as follows: Under the stabilizing influence of rotation, convection cells in a fluid that is initially isothermal and finally approximately critical in vertical lapse rate occurred with their central cores moving away from the horizontal surface near which

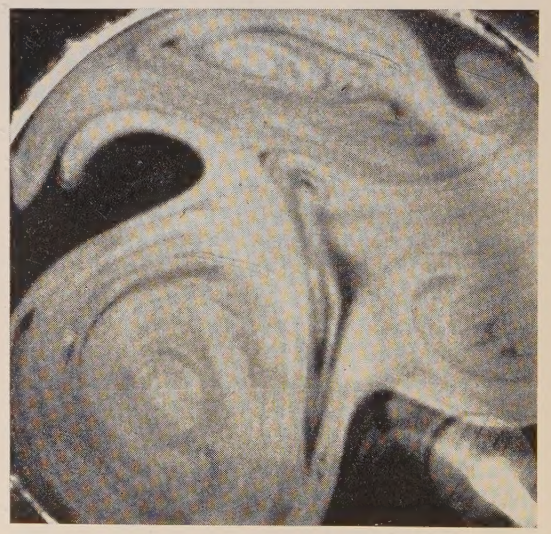


Fig. 6. Top view of a cross section 37 cm from the bottom of a cylinder of air with conditions similar to those of Fig. 5. The smoke tracer is diverging and rotating anticyclonically in the upper portions of the ascending cell cores. (Approx. one third size.)

the lapse rate had first attained the critical value. As we shall see, the theory suggests that the role of the rotation in this process is to increase the required critical lapse rate and to make it improbable that it be first attained anywhere other than near a surface where heat is being actively added or removed.

It is to be noted that this statement applies to lapse rate changes as they have been observed to occur in the experiments and not as they are assumed to occur in the theoretical discussions, below. With the relatively high rates of heating or cooling of the experiments, the critical lapse rate builds up or down from the appropriate surface, leaving the remaining isothermal layers undisturbed. In the theoretical analysis the lapse rate is assumed to increase uniformly, maintaining a linear character and finally establishing the critical lapse rate simultaneously throughout the depth of the fluid. Experimentally, with a low enough heating rate, adequate side insulation, and a provision to remove heat from the top at the same slow rate, this linear assumption might be fulfilled. If all the levels of an undisturbed fluid layer were brought to instability simultaneously in this ideal manner, the direction of the cells' vertical circulation would very probably either be completely arbitrary, or would be determined by some smaller effect such as the variation of viscosity with temperature,

Theoretical Analysis

As Jeffreys and others have pointed out, the effect of a basic rotation upon convection is a stabilizing one. There are two other, less apparent effects: 1) a special "overstable" oscillation appears as the first mode of instability in certain fluids under large rotational influence; and 2) the diameters of convection cells vary inversely with the rate of rotation, the ratio of the depth to the diameters of cells increasing markedly. The effects can be shown by extending Rayleigh's treatment to rotating systems.

For the analysis, the following equations apply.

The Eulerian form of the Navier-Stokes equations of motion for a fluid system rotating around the vertical z axis with angular velocity Ω:

$$\frac{Du}{Dt} - 2\Omega v = -\frac{1}{\varrho}\frac{\partial p}{\partial x} + \nu\nabla^2 u + \Omega^2 x + \frac{\nu}{3}\frac{\partial X}{\partial x} \quad (1)$$

$$\frac{Dv}{Dt} + 2\Omega u = -\frac{1}{\varrho}\frac{\partial p}{\partial y} + \nu\nabla^2 v + \Omega^2 y + \frac{\nu}{3}\frac{\partial X}{\partial y} \quad (2)$$

$$\frac{Dw}{Dt} = -\frac{1}{\varrho}\frac{\partial p}{\partial z} + \nu\nabla^2 w - g + \frac{\nu}{3}\frac{\partial X}{\partial z} \quad (3)$$

where ν is the kinematic viscosity,

$$X = \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right),$$

and the other symbols have their familiar meanings.

The equation of continuity:

$$\frac{D\varrho}{Dt} + \varrho(X) = 0 \quad (4)$$

The energy equation (GOLDSTEIN, 1943):

$$\frac{D\theta}{Dt} + \frac{p}{\varrho J c_v}(X) = \varkappa\nabla^2\theta + \frac{\Phi}{\varrho J c_v} \quad (5)$$

where θ is the temperature, $\varkappa$ the thermometric conductivity, J the mechanical equivalent of heat, and Φ is the "dissipation function". The equation of thermal expansion:

$$\varrho = \varrho_1(1 - \alpha\triangle\theta) \quad (6)$$

where $\Delta\theta$ is the difference between the temperature at which the density is ϱ and the temperature at which the density is ϱ_1; α is the coefficient of thermal expansion.

Superposing small perturbations upon an initial state of relative rest (solid rotation) and neglecting terms of higher order involving the products of small quantities, a set of perturbation equations of motion is obtained. These are further simplified through a special incompressibility assumption first applied to the cellular convection problem by Rayleigh[1]: the fluid is considered incompressible, the density constant, except in so far as thermal expansion affects the vertical equation of motion through the gravity term. In addition, the horizontal forces arising from centrifugal accelerations are considered small enough to be neglected. This essentially amounts to regarding the expansion coefficient α as a small quantity of the first order in the perturbation analysis. With

[1] Rayleigh credits the introduction of this assumption to BOUSSINESQ, *Theorie Analytique de la Chaleur*, vol. II, p. 172 (1903).

these considerations the perturbation equations of motion reduce to

$$\frac{\partial u}{\partial t} - 2\Omega v = -\frac{1}{\varrho_0}\frac{\partial p}{\partial x} + \nu\nabla^2 u \tag{7}$$

$$\frac{\partial v}{\partial t} + 2\Omega u = -\frac{1}{\varrho_0}\frac{\partial p}{\partial y} + \nu\nabla^2 v \tag{8}$$

$$\frac{\partial w}{\partial t} = -\frac{1}{\varrho_0}\frac{\partial p}{\partial z} + \nu\nabla^2 w + \alpha\theta g \tag{9}$$

where the usual primes have been omitted and u, v, w, p, and θ now represent perturbation quantities.

Under the Rayleigh-Boussinesq assumption, the perturbation equation of continuity becomes

$$X = 0 \tag{10}$$

where we have again treated α as a small quantity of the first order.

Assuming an initial state of relative rest, terms of the energy equation (5) involving velocity components or their derivatives disappear. Assuming in addition an initial steady state temperature distribution (conductive equilibrium) and considering the temperature field uniform in the horizontal, we have

$$\theta_0 = \Theta - \beta z \tag{11}$$

where $\beta = -\dfrac{\partial\theta_0}{\partial z}$ and Θ is the temperature at $z = 0$. Substituting (11) in the reduced energy equation and applying the usual perturbation method we obtain

$$\frac{\partial\theta}{\partial t} - \beta w = \varkappa\nabla^2\theta \tag{12}$$

for the perturbation energy equation.

Equations (7), (8), (9), (10), and (12) are solved simultaneously to eliminate u, v, θ, and p.[1] An eighth order partial differential equation in w results:

$$\left(\frac{\partial}{\partial t} - \varkappa\nabla^2\right)\left(\frac{\partial}{\partial t} - \nu\nabla^2\right)^2\nabla^2 w - \alpha\beta g\left(\frac{\partial}{\partial t} - \nu\nabla^2\right)\cdot \left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right) + 4\Omega^2\left(\frac{\partial}{\partial t} - \varkappa\nabla^2\right)\frac{\partial^2 w}{\partial z^2} = 0 \tag{13}$$

[1] If u, v, w, and p are eliminated, a similar eighth order expression results, but for θ in place of w. Thus the field of perturbation vertical velocity and the perturbation temperature are of similar configuration.

The solution of (13) is assumed to be separable in the form

$$w = f(x, y)\, g(z)\, h(t)$$

and is required to meet the following conditions: 1) it must have a cellular distribution in the horizontal, i.e., $\nabla_h^2 f(x,y) \propto f(x,y)$; 2) it must fit the boundary conditions appropriate to free surfaces at the top and bottom, i.e., $w = 0$, $\theta = 0$, $\partial\zeta/\partial z = 0$, and $\partial^2 w/\partial z^2 = 0$. A form meeting these requirements is

$$\left.\begin{aligned} w = W_0 \cos(k_x x) \cos(k_y y) \\ \cos[(\pi/D)z] \exp(nt) \end{aligned}\right\} \tag{14}$$

where $k_j = 2\pi/L_j$, the wave number in the direction of j with corresponding wave length L_j, for $j = x,y$; W_0 is an arbitrary constant. The solution represents a rectangular cell of horizontal dimensions L_x and L_y with free top ($z = +D/2$) and bottom ($z = -D/2$) surfaces. For stability considerations, a time dependent exponential factor has been included in the usual manner.

Substituting (14) in (13) we obtain a third degree frequency equation for n:

$$n^3 + A_1 n^2 + A_2 n + A_3 = 0 \tag{15}$$

where

$$A_1 = \frac{(\varkappa + 2\nu)(a^2 + \pi^2)}{D^2} \tag{16}$$

$$A_2 = \frac{\nu(2\varkappa + \nu)(a^2 + \pi^2)^2}{D^4} - \frac{\alpha\beta g a^2}{(a^2 + \pi^2)} + \frac{4\Omega^2\pi^2}{(a^2 + \pi^2)} \tag{17}$$

$$A_3 = \frac{\varkappa\nu^2(a^2 + \pi^2)^3}{D^6} - \frac{\alpha\beta g a^2 \nu}{D^2} + \frac{4\Omega^2\pi^2\varkappa}{D^2} \tag{18}$$

in which

$$a^2 = D^2(k_x^2 + k_y^2) \tag{19}$$

the two dimensional wave number (cf. PELLEW and SOUTHWELL, 1940). The roots, n, of this frequency equation, appearing as they do in the time dependent factor $\exp(nt)$ in (14), determine the stability condition of the system. The possible roots of the cubic, the criteria for their appearance in terms of the coefficients of the frequency equation (cf. VON KARMAN and BIOT, 1940), and the physical interpretation of

Table 1

Real root n	$n < 0$	$A_3 > 0$	damped; perturbation decreases	
	$n = 0$	$A_3 = 0$	neutral; perturbation steady	
	$n > 0$	$A_3 < 0$	amplified; perturbation increases; ordinary convective instability	
Complex root $m + ib$	$m < 0$	$A_1A_2 - A_3 > 0$	perturbation causes oscillation	damped; decreasing amplitude
	$m = 0$	$A_1A_2 - A_3 = 0$		neutral; steady amplitude
	$m > 0$	$A_1A_2 - A_3 < 0$		amplified; increasing amplitude; "overstable" oscillation

Criteria for roots of the frequency equation and the corresponding physical behavior of the perturbation. (Provided $A_1 > 0$.)

the behavior of the imposed perturbation are summarized in Table 1. Two kinds of instability are indicated, cells which grow monotonically and cells which oscillate with increasing amplitude.

It is assumed that a system initially stable and proceeding toward instability will pass through a state of neutral equilibrium on the way (the principle of "marginal stability"). Therefore, $A_3 = 0$ or $A_1A_2 - A_3 = 0$ will preface the appearance of one or the other kind of instability. An evaluation of each of these criteria shows the range of conditions for the first appearance of either kind in preference to the other.

Ordinary Convective Instability

Rewriting the condition $A_3 = 0$ from (18) we obtain

$$(a^2 + \pi^2)^3 - Ra^2 + T\pi^2 = 0 \qquad (20)$$

where the Rayleigh number

$$R = \frac{\alpha\beta g}{\varkappa\nu} D^4 \qquad (21)$$

and the Taylor number[1]

$$T = \frac{4\Omega^2}{\nu^2} D^4 \qquad (22)$$

Since $R = f(a^2, T)$, for a specific T we may write $R = f(a^2)$. Fig. 7 plots three such relations (heavy solid line) for selected Taylor numbers. At any particular value of T there is a minimum point designating the first occurrence of instability in a system proceeding from stability; the magnitudes of both the minimum Rayleigh number for convection and the wave number of the cells that first occur increase with increasing values of T.

From (20) we find $\partial R/\partial(a^2)$ and set it equal to zero to obtain

$$R_c = 3\,(a_c^2 + \pi^2)^2 \qquad (23)$$

where the subscript c refers to the critical value of these parameters. Substituting (23) for R in (20) we have

$$(a_c^2 + \pi^2)^3 - 3\,(a_c^2 + \pi^2)^2 a_c^2 + T\pi^2 = 0 \qquad (24)$$

The only real root of this equation is

$$a_c^2 = \pi^2 \cosh\frac{\phi}{3} - \frac{\pi^2}{2} \qquad (25)$$

where

$$\phi = \cosh^{-1}\left(1 + \frac{2T}{\pi^4}\right) \qquad (26)$$

By substituting (25) in (23), the relation between T and R_c for ordinary convective instability is obtained:

$$R_c = 3\pi^4\left(\frac{1}{4} + \cosh\frac{\phi}{3} + \cosh^2\frac{\phi}{3}\right) \qquad (27)$$

This expression is plotted in Fig. 8 (heavy solid line). Beyond $T = 10^4$ the curve approximately parallels $R_c = T^{2/3}$ on a log-log plot. At $T = 0$, $R_c = 27\pi^4/4$ and $a_c^2 = \pi^2/2$, values which are identical to those obtained by others for the non-rotating, two free boundary case.

[1] This quantity, originally used in this particular form by Chandrasekhar readily derives from a parameter first introduced by G. I. Taylor.

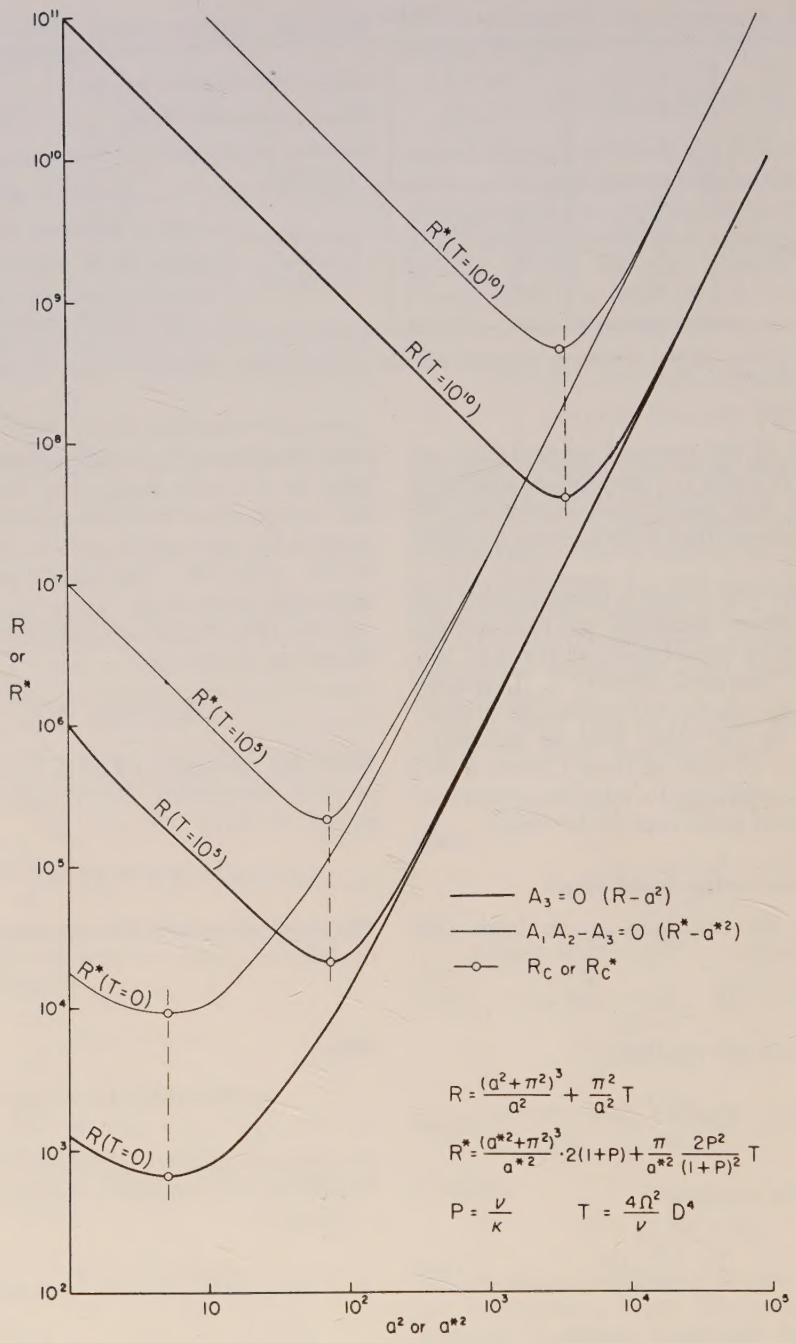


Fig. 7. Wave number versus required Rayleigh number for ordinary convective and overstable oscillating modes (starred quantities) in water ($P=6$) at several values of the Taylor number. Note minimum Rayleigh number for a specified Taylor number, designating the conditions for the first appearance of instability.

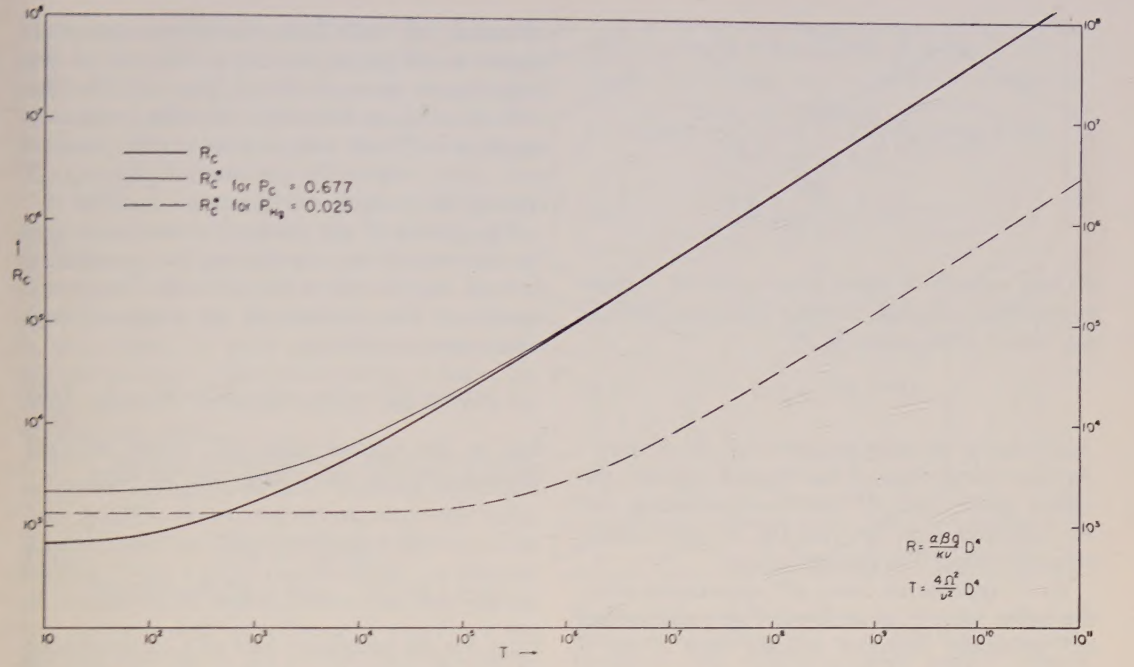


Fig. 8. Taylor number versus critical Rayleigh number for the first appearance of ordinary convection (R_c), for overstable oscillation (R_c^*) at the critical Prandtl number ($P=0.677$), and for overstable oscillation at the Prandtl number of mercury ($P=0.025$).

Overstable Oscillation

Consider now the second condition, $A_1A_2 - A_3 = 0$. Substituting for the coefficients from (16), (17), and (18) we have

$$2(1+P)^2(a^2+\pi^2)^3 - Ra^2(1+P) + 2P^2T\pi^2 = 0 \tag{28}$$

where the Prandtl number

$$P = \nu/\varkappa \tag{29}$$

By operations similar to those used to find R_c from (20) we obtain

$$a_c^{*2} = \pi^2 \cosh \frac{\varphi^*}{3} - \frac{\pi^2}{2} \tag{30}$$

$$R_c^* = 6(1+P)\pi^4 \left(\frac{1}{4} + \cosh\frac{\phi^*}{3} + \cosh^2\frac{\phi^*}{3}\right) \tag{31}$$

where $\phi^* = \cosh^{-1}\left[1 + \left(\frac{P}{1+P}\right)^2 \frac{2T}{\pi^4}\right]$ (32)

Since R_c^* is a function of both T and P, it is represented by a family of curves rather than a single one as was possible with R_c. Two such curves are plotted in Fig. 8 for two values of P (thin solid and thin broken lines).

These starred quantities, deriving as they do from the criterion for the occurrence in the frequency equation of complex roots with positive real parts, are appropriate to the first appearance of the second kind of instability, oscillations of increasing amplitude. Chandrasekhar calls the phenomenon "overstability" after Eddington who developed the concept in a theory concerning the pulsations of Cepheid variable stars. Eddington's choice of the term is best explained in his own words: "In the usual kind of instability a slight displacement provokes forces tending away from equilibrium; in overstability it provokes restoring forces so strong as to overshoot the corresponding position on the other side of equilibrium and sets up an increasing oscillation" (EDDINGTON, 1926, p. 201).

In the region of larger values of T, members of the R_c^* family also parallel a $T^{2/3}$ curve. Specifically, for increasing values of T, R_c and R_c^* approach the respective asymptotes

$$R_c = \frac{3}{4^{1/3}}\pi^{4/3}T^{2/3} \cong 8.7\ T^{2/3} \qquad (33)$$

$$R_c^\star = 6(1+P)\left(\frac{\pi^4}{4}\right)^{1/3}\left(\frac{P}{1+P}\right)^{4/3}T^{2/3} \cong$$

$$\cong 17.4 \cdot \frac{P^{4/3}}{(1+P)^{1/3}}T^{2/3} \qquad (34)$$

At one value of P and a large value of T these asymptotes coincide. Setting (33) equal to (34) we obtain a biquadric in P:

$$8P^4 - P - 1 = 0 \qquad (35)$$

which has as its only positive root, $P_c \cong 0.677$.[1] At this critical value of the Prandtl number, the values of R_c and $R_c^\star$ practically coincide for all T greater than 10^6 (see Fig. 8) and exactly coincide when T is infinitely large.

As T approaches zero, $a_c^{\star 2}$ approaches $\pi^2/2$, the same value as a_c^2 at $T=0$. Fig. 7 shows that $a_c^{\star 2}$ gradually becomes smaller than a_c^2 as T increases. Thus, if both types of convection were occurring at the same non-zero value of T, the overstable cells would be somewhat larger in diameter than the ordinary type. This point leads to an interesting meteorological application in a later section.

Also as T approaches zero, $R_c^\star$ approaches $27\frac{\pi^4}{2}(1+P) \cong 1315\,(1+P)$, a value which is never smaller than twice as large as R_c for the non-rotating case. This would seem to imply that complex roots with positive real parts might exist in the case of non-rotation, provided only that the Rayleigh number be made large enough. But RAYLEIGH, JEFFREYS (1926), and PELLEW and SOUTHWELL showed that such roots could only occur with negative real parts in the non-rotating case. The point of departure of these results can be traced to the partial differential equation (13). If from this point one proceeds on the assumption that $\Omega = 0$, then the third term on the left disappears and each of the remaining terms contain as a common factor the operator $(\partial/\partial t - \nu\nabla^2)$ which can be cancelled. This reduces the order of the partial differential equation by two and the degree of the frequency equation by one. But the quadratic frequency equation thus obtained can only have complex roots with negative real parts, for the coefficient of the first degree term is always positive. On the other hand, as we have seen, the cubic frequency equation will have complex roots with positive real parts when the relation $A_1A_2 - A_3 < 0$ among the positive coefficients is fulfilled.

The period of the overstable oscillation can be determined by considering the imaginary part of the complex roots of the frequency equation. The product of the roots of (15) taken two at a time is

$$n(m+ib) + n(m-ib) + m^2 + b^2 = A_2 \qquad (36)$$

But at the critical value $R_c^\star$, $m=0$; so (36) becomes $b^2 = A_2$. Rewriting (17) we have

$$\left.\begin{aligned} b^2 = \frac{\varkappa^2}{D^4}\frac{P}{(a_c^{\star 2}+\pi^2)} \cdot \\ \cdot\left[(2+P)(a_c^{\star 2}+\pi^2)^3 - R_c^\star a_c^{\star 2} + PT\pi^2\right] \end{aligned}\right\} \qquad (37)$$

Since b is the imaginary part of a complex root of the frequency equation, the period of oscillation is

$$\tau = \frac{2\pi}{b} \qquad (38)$$

Summary of Instability Criteria

So the question of which type of instability appears before the other is decided by an intrinsic characteristic of the fluid medium, the Prandtl number, and the rotational parameter, the Taylor number. For fluids with $P > 0.677$ (e.g., water, air), R_c is smaller than $R_c^\star$ for all T, and therefore ordinary convective stability will always occur first. For fluids with $P < 0.677$ (e.g., mercury), the critical Rayleigh curves for the two types of instability cross at some point, $R_c^\star$ becoming the smaller in regions of large T (see Fig. 8). Thus, ordinary convection would be the first to occur at low values of T, and the overstable oscillating mode would take priority at large values of T. In the case of fluids with $P \geq P_c$, the R_c curve alone describes a general criterion for the breakdown of stability; in fluids with $P < P_c$, the appropriate critical curve would be the envelope containing the R_c curve from $T=0$ to the crossover point (a function of P) and the $R_c^\star$ curve from the crossover to infinitely large T.

The curves plainly show the stabilizing effect of rotation first pointed out by Jeffreys, R_c

[1] An equivalent value was obtained by Chandrasekhar, though he writes the reciprocal, $\kappa/\nu = 1.478$.

increasing rapidly beyond $T = 10^2$ for ordinary convection, and beyond some $T(P)$ for overstable oscillation. Jeffreys' observation that high viscosities could actually aid instability can also be seen. Since $T \propto \nu^{-1/3}$ and $R \propto T^{3/3}$ for large T, it readily follows that $\beta \propto \nu^{-1/3}$.

It has been estimated that high conductivity might also assist instability by "preventing the (stabilizing) effects of rotation from being dominant", but this does not follow since $\varkappa$ appears only in the denominator of R_c. There is, as it were, no early need for a convective regime to set in if the conductivity is high, for large amounts of heat can be transferred by conduction alone.

The validity of Stommel's limit, rotational influence negligible in fluid layers of depth less than $(\nu/\Omega)^{1/2}$, becomes evident when this value is substituted for D in the Taylor number. The result is $T = 4$, a point clearly in the flat portion of the curve, R_c having increased only about one per cent over its value at $T = 0$. For a ten per cent increase in the critical Rayleigh number, the limit becomes $D \leq 1.7\,(\nu/\Omega)^{1/2}$; for a one hundred per cent increase, the constant multiple is 3.3.

Cell Shapes

Experimental observations have shown that cells of hexagonal shape most often appear (centers of adjacent cells are usually located near the vortices of an equilateral triangle, cf. Figs 1 and 10). Beginning with an expression for w representing a hexagonal distribution (Christopherson, 1940),

$$w = \frac{W_0}{2}\left[\cos\frac{2\pi}{3L}(3^{1/2}x + y) + \cos\frac{2\pi}{3L}(3^{1/2}x - y) + \right.$$
$$\left. + \cos\frac{4\pi}{3L}y\right]\cos\frac{\pi}{D}z \qquad (39)$$

where W_0 is an arbitrary constant and L is the length of one side, an expression for the streamlines at the top surface can be derived; for large T,

$$\Psi = \left(\cos\frac{2\pi}{3L}3^{1/2}x\right)\left(\cos\frac{2\pi}{3L}y\right) + \frac{1}{2}\cos\frac{4\pi}{3L}y \qquad (40)$$

Because of the restriction to large T (considering negligibly small those terms which do not contain T), equation (40) represents concentric circles instead of spirals. These streamlines and a set of relative vorticity values are plotted in Fig. 9.

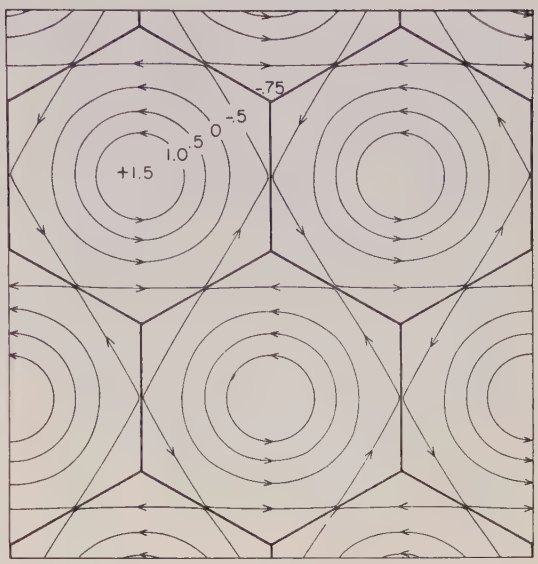


Fig. 9. Theoretical streamlines for the top surface of hexagonal convection cells at a high Taylor number. Numerical data represent arbitrary units of relative vorticity.

The effect of rotation upon the overall shapes of cells, i.e., the increase of the cells' depth-diameter ratio, can be demonstrated by a further consideration of the two dimensional wave number. From equation (39)

$$a^2 = D^2\left(\frac{4\pi}{3L}\right)^2 \qquad (41)$$

Combining this with the distance between the centers of adjacent hexagons as an average cell diameter, we obtain for the depth-diameter ratio

$$\frac{D}{d} = \frac{3^{1/2}}{4\pi}a_c \qquad (42)$$

for which a_c may be determined from (25). The increase in depth-diameter ratios with increasing rates of rotation can be seen in Table 2.

Table 2

$T =$	0	10^3	10^6	10^9	10^{12}
$D/d =$	0.3	0.5	1.8	5.7	18.0

The increase in cellular depth-diameter ratios with increasing Taylor number in water at 25° C.

Table 3

	D/d	
	Predicted	Measured
Fig. 1	2.1	2.1
Fig. 2	14.3	15.5
Fig. 10	5.8	5.6

Cellular depth-diameter ratios predicted theoretically compared to those measured in several of the illustrated experiments.

Table 3 lists the ratios predicted by (42) for several of the experiments illustrated and compares them with actual measurements.

The Experimental Apparatus

The program of experiments intended to test the theoretical results was designed with the non-dimensional Rayleigh and Taylor parameters in mind. Water was used as both a readily available medium and one whose physical properties are accurately known and tabulated (Dorsey, 1940). It remained for the experimenter to measure the angular velocity of the rotating model and the depth, mean temperature, and vertical lapse rate of the fluid. By changing the depth and/or the angular velocity a range of experiments at different values of the Taylor number was carried out, measuring the lapse rate at which convection began in each one as a dependent variable. In this way, an array of points along the R_c, T curve was determined.

For steady rotation of the fluid models a leveled, eighteen inch diameter, three inch wide steel ring was mounted on a vertical shaft. The ring provided a load bearing surface with provision for leveling, centering, and clamping various experimental models in place. The shaft was driven through V-belts by a constant speed induction motor with a continuously variable transmission providing a range of speeds at the mounting ring of from zero to fifty revolutions per minute. An electric counter-timer circuit measured the time required for any whole number of revolutions to one hundredth second accuracy.

Containers for the experiments were made from open end pyrex glass cylinders of twelve inch nominal outside diameter. Pyrex bottoms were cemented to one end of each cylinder, electrically conducting pyrex heating plates for experiments heating from below and ordinary pyrex plate for those with cooling from above.

Power for the heating plates was supplied by a variable transformer and was carried into the rotating system through mercury slip-rings. A wattmeter measured the input; but due to extraneous losses, only a fraction of the indicated input was used to drive the cells, so the meter readings could only be used comparatively. Cooling from above was usually accomplished by evaporation at the free surface, but a few trials were made with an external heat sink. A second, thin walled cylinder of slightly smaller diameter was mounted in the upper part of the main cylinder, its bottom in contact with the water surface. Cool water from a thermostatically controlled mixing valve was then continuously circulated through the upper cylinder, making it an approximately constant heat sink of adjustable capacity.

Visual observations were made with the aid of a rotoscope, a reversing prism mounted with its optical axis along the axis of rotation of the turntable and rotated at exactly half the table's speed by a selsyn drive geared to the main shaft. With the aid of this device, all motions in the model are seen relative to an apparently stationary turntable.

Top surface circulations were made visible by small tracer particles of aluminum powder, their movements either viewed directly with the aid of the rotoscope or photographed with the time exposure, "streak" technique. An electronic flash triggered just before the end of an exposure marked the direction in which the particles were travelling. Fig. 10 is a positive reversal of one such "streak-flash" photo. Evaporation cooling was driving cells with descending cores, and although there is some confusion where streaks overlap and where spurious flashes appear because of reflections from the particles, the majority of flashes indicate that the circulation was cyclonic (southern hemisphere sense, since the basic rotation was clockwise).

Internal cell motions were traced with nigrosine ink and photographed by conventional instantaneous flash technique (Figs 3 and 4). To illuminate the smoke tracer (ammonium chloride) circulating in cells in air, the light from the electronic flash was collimated by a

horizontal slit and projected through the experimental cylinder at a selected level while the camera viewed the cross section from above (Figs 5 and 6).

The vertical temperature gradient was the crucial variable of the experiment. It was evident from the theoretical results that the lapse rates would be so very small that a sensitive measuring system would be required. The critical Rayleigh number for ten centimeters of water rotating at ten revolutions per minute calls for a lapse rate of only 0.025° C per cm. Further, the temperature sensitive elements would have to be small enough so as not to interfere with the delicate cellular circulations, and they would have to be able to transmit their information to a measuring device outside the rotating system.

These requirements were met with copper-constantan wire thermocouples. Ten junctions connected in a series thermopile gave an output of a little better than two hundred microvolts per degree Centigrade and also provided a five point sample of the temperature at each level, thus serving to average out the horizontal temperature gradients within the cells. The mean temperature of the fluid was taken at the end of each run with a mercury thermometer and used to determine the tabulated values of α, $\varkappa$, and ν.

The temperature signal was carried from the model to the rotating member of a set of mercury slip-rings, six concentric copper cylinders mounted in a lucite plate that turned with the main shaft. The stator consisted of a thicker lucite plate in which six concentric circular troughs had been cut, each then filled to about two-thirds of its depth with mercury. The copper cylinders rotated freely in the stator troughs with their lower edges well under the mercury surface. The device maintained three virtually constant resistance circuits without contact "noise" between the rotating and non-rotating parts of the apparatus. The mercury slip-rings introduced little or no detectable errors into the thermopile circuit, for there was no opportunity for a temperature difference to build up across the thermocouple consisting of the copper rotor, the continuously mixed mercury, and the copper terminals of the stator.

From the mercury troughs, the signal was carried to an automatic potentiometer that balanced the small voltage input and provided a proportional current which was fed in turn to a recording milliammeter. The entire system, used in conjunction with a ten junction thermopile, was capable of detecting temperature differences reproducible within ±0.01° C.

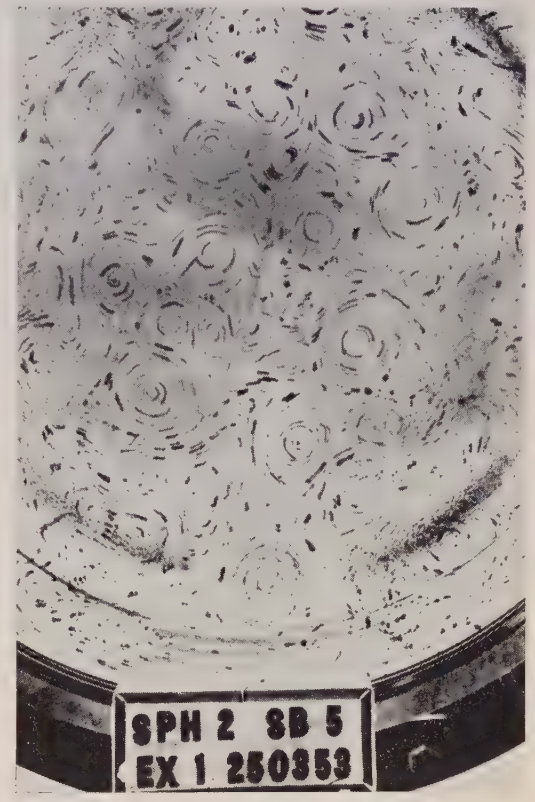


Fig. 10. A positive reversal of a streak photograph of the top surface of cells driven by evaporation cooling in water. Water 22.6° C; air 23.3° C; depth 18 cm; rotating 5.0 rpm; Taylor number 1.2×10^9; exposure time four seconds. (Approx. one half size.)

Experimental Verification of the R_c Criterion

It was evident that lapse rate measurements should be made with as steady conditions as possible in order best to approximate the assumptions of the theoretical analysis. It seemed that this could be done by supplying only the minimum amount of heating from below necessary to drive the convective circulation. This was investigated by studying the character of vertical temperature difference records of the type illustrated in Fig. 11. In the curve at the top, the heating rate is too low to produce convective instability; the vertical temperature

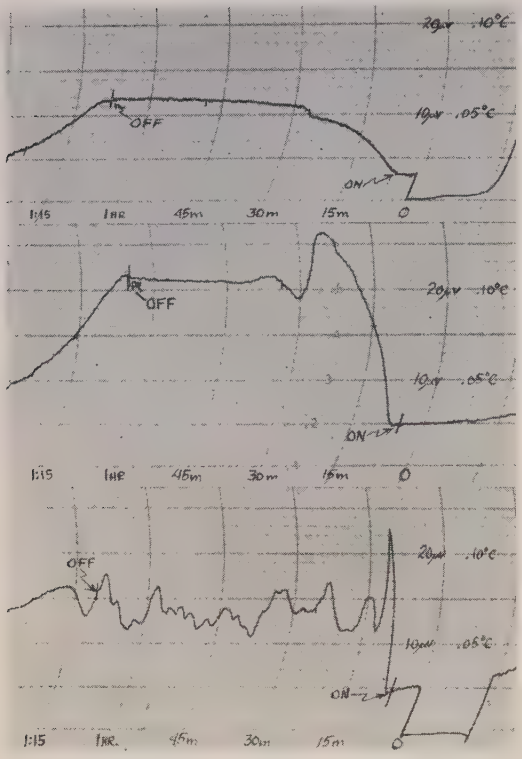


Fig. 11. Vertical temperature difference records made at three different rates of heating from below. Depth 3 cm; rotating 10 rpm; thermopiles at 1.0 and 1.9 cm levels; one, three, and ten watts heating plate input, top to bottom.

difference trace rises gradually and asymptotically approaches some constant value where the vertical transfer of heat by conduction alone is balanced by the losses at the top surface. In the bottom curve, the heating rate is too large for steady convective heat transfer; the temperature trace rises sharply, breaks down suddenly as unsteady cells are established, and continues to oscillate with the changing vertical circulations. The middle curve is characteristic of a heating rate somewhat greater than the minimum, in this case, three watts, whereas the minimum was something nearer two watts. The trace rises smoothly to a supercritical peak value, breaks down as convection begins, and goes into a "damped oscillation" that levels out at the critical lapse rate value after the initial overturning is completed and approximately steady state convection has been fully established. In the experimental series, greater than minimum heating rates were used, since with bare minimums it became apparent it would take hours to establish cells. In general, rates that could establish cells in from fifteen to twenty minutes were found convenient; visual observations of the cellular circulations under these conditions showed them to be as steady as could be desired.

A series of experiments were carried out in which the mean temperature at five different levels was measured within thirty seconds by switching the recording system across five different thermopiles. These showed that the lapse rate became essentially linear in the middle layers of the fluid ten or fifteen minutes after convection had begun, but that it was steeper than average within both the lowest centimeter where heat was being added and in the top two or three centimeters near the evaporation cooled free surface. Excess cooling was reduced by partially restricting the ventilation of the top surface with a porous lid. In all subsequent experiments, average lapse rates were measured in the middle layers of the fluid. The thermopile grids were set as far apart vertically as was possible in the particular fluid depth, but they were kept clear of the excessively warm lower and cold upper layers.

In Fig. 12 the values of critical Rayleigh numbers obtained in these experiments are plotted along a section of the theoretical R_c, T curve. The close correspondence between the experimental points and the theoretical curve agrees with Chandrasekhar's result that the two free and the one rigid, one free boundary cases are closely equivalent except when the influence of rotation is low, for the curve was derived under the assumption of two free boundaries whereas the points were observed in a model in which one boundary was rigid.

Experimental Verification of the R_c^* Criterion

At the time of the present preliminary report, experiments designed to investigate the overstable mode have begun. Mercury, with a Prandtl number of 0.025, well under the critical 0.677, is the fluid used in these trials. The dashed curve of Fig. 8 is the theoretical relationship for the occurrence of the second kind of instability in mercury, again under the assumption of two free bounding surfaces. For all values of the Taylor number in excess of 10^3, R_c^* is less than R_c and therefore the overstable mode will be the first to occur.

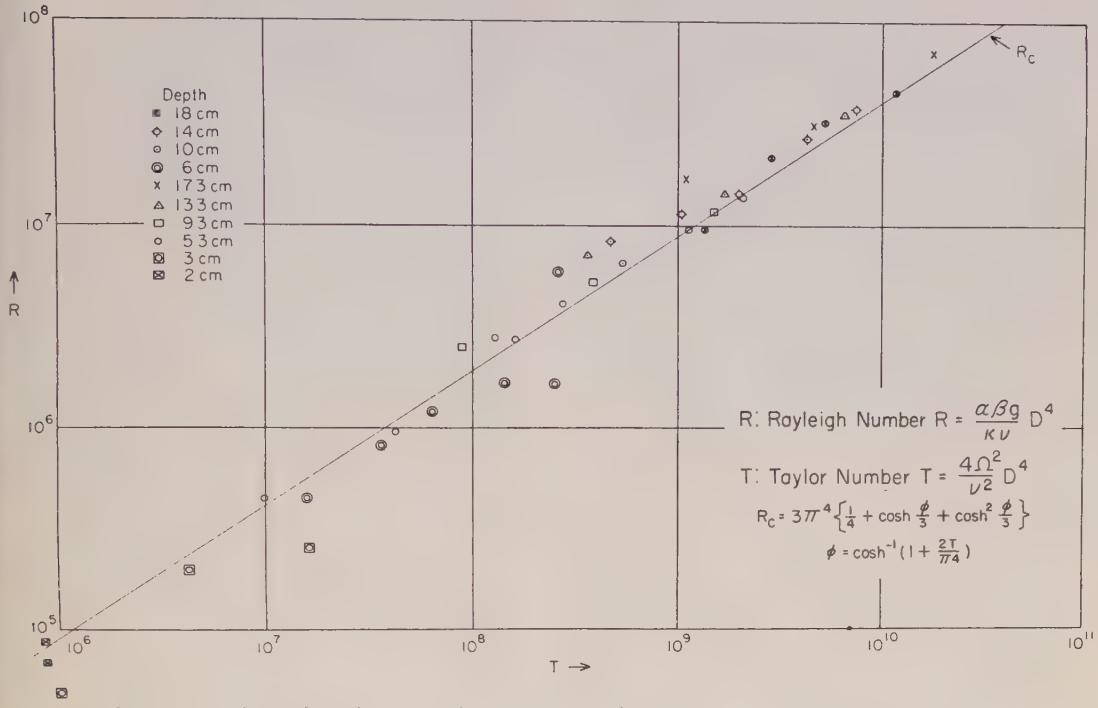


Fig. 12. Experimental results versus theoretical curve for ordinary cellular convection in water.

The first trials in this series have shown that the phenomenon of overstability does exist (Fultz, Nakagawa & Frenzen, 1954). A five centimeter deep layer of mercury was heated from below while rotating approximately ten revolutions per minute. Several cell-like circulations, each about two centimeters in diameter, appeared at the top surface, their motion made visible by small particles of impurities. Each cell periodically reversed the direction of its relative rotation, implying alternating divergence and convergence at the top surface and hence alternating ascent and descent in the central core. The period of the oscillation was about fifteen seconds, very close to the 13.9 seconds predicted by equation (38).

Recalling Eddington's definition of overstability, ". . . a slight displacement . . . provokes restoring forces so strong as to overshoot the corresponding position on the other side of equilibrium", one can describe a mechanism for these oscillating cells. If in a system heated from below we have a fluid of sufficiently small Prandtl number, i.e., a fluid with a high enough thermometric conductivity and low enough kinematic viscosity, then the lower layers readily absorb heat and, expanding, soon become sufficiently buoyant to overcome the restraint imposed by the viscosity. Convective cells begin to form in the usual manner, the cores carrying the hot fluid upward and the cell walls bringing colder fluid downward to replenish the lower layers. Because of its high conductivity and a consequent ability to absorb an excess amount of heat before vertical motion can begin, the fluid attains more buoyancy than that required merely to overcome the weak viscous stresses. Therefore the overturning takes place rather more violently than is the case with steady, ordinary convection cells. The rapid transfer of hot fluid upward and cold fluid downward destroys the unstable density stratification, and the fluid begins to deccelerate. But since the fluid began its motion with an excess buoyancy, it initially "overshoots the corresponding position on the other side of equilibrium". Excess cooling in the upper layers, again made possible by the high conductivity, drives the other half of the cycle in a similar manner, but in the opposite direction. In practice, the fluid will not continue to overshoot further and further at the end of

each swing, since there are practical limits to the fluid velocities and hence to the amplitude of oscillation. In short, during the oscillation the fluid alternates between two states of unstable stratification, one the equivalent of super-critical heating from below and the other, of super-critical cooling from above.

In a rotating system the tendency for fluid rings to resist horizontal divergence supplies an additional restoring force when the rings are forced to move away from their equilibrium position by the appearance and disappearance of fluid in their centers. It may be that thermal expansion effects alone are insufficient to maintain overstable oscillations, and that for this reason the occurrence of this mode is not physically possible in non-rotating systems. Since this additional restoring force is a result of rotation, this is at least cosistent with the theoretical result obtained by the inclusion of a non zero rotation in the frequency equation, i.e., the increase of the equation's degree which makes possible the appearance of complex roots with positive real parts and their associated overstable mode.

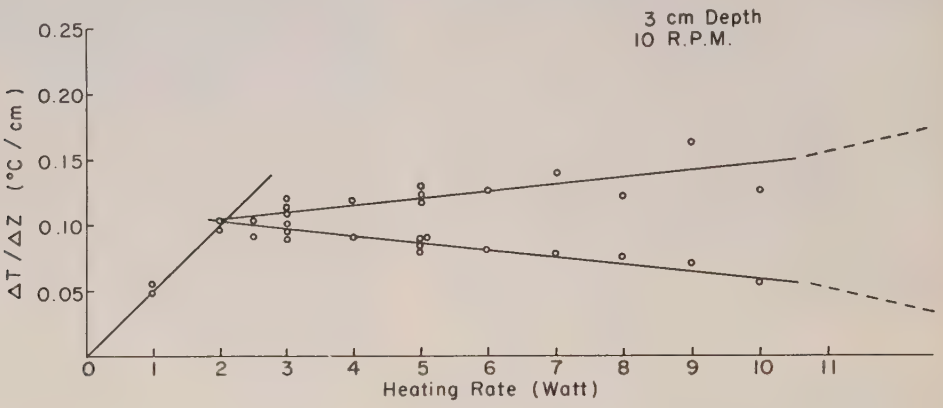


Fig. 13. Heating rate versus characteristic lapse rate values. Between 0 and 2 watts, the curve represents the maximum steady lapse rate attained; beyond 2 watts the upper branch represents the maximum supercritical lapse rate attained, and the lower branch represents the value at which the lapse rate steadies or around which it oscillates after convection has begun.

Regimes of Vertical Heat Transfer

The experiments of the type illustrated in Fig. 11, those involving the changing character of the temperature trace with changing rates of heating, were first carried out merely with the idea of determining the minimum heat required for convection. But they can also be interpreted from the standpoint of the several modes of heat transfer involved. The trials illustrated were carried out in a three centimeter deep layer of water, rotating at ten revolutions per minute. The temperature values recorded are the differences between the 1.0 and 1.9 centimeter levels. The only change in the experiment between the separate traces is in the heating rate.

When the power supplied is one watt, heat is transferred vertically by conduction alone. Approximately steady state convective heat transfer occurs when the system is heated with three watts input at the plate, while ten watts input leads to an unsteady convective transfer regime. An interesting relation becomes apparent when characteristic temperatures from traces obtained in experiments of this type are plotted against their heating rates (Fig. 13). In the conductive regime (0 to 2+ watts), the heating rate is plotted against the steady, maximum temperature difference attained. At heating rates greater than the minimum required for convection, we plot two characteristic temperature differences, the amplitude of the first peak which breaks down when convection begins, and the average value at which the trace later steadies or around which it oscillates.

The upper branch of Fig. 13, the plot of the supercritical peak attained before convection, uniformly increases with increasing heating rates up to regions in which the breakdown occurs sharply, almost immediately after the

heat is turned on. The lower branch, representing a value proportional to the average vertical temperature difference that can be maintained with the prevailing rates of heat input and vertical transfer, decreases continuously in an approximately linear fashion over the range of relative small heating rates covered by Fig. 13. The triple point located by the intersection of the three curves designates the critical lapse rate and the minimum heating rate required for convection. The discontinuous slope consisting of the conductive and the steady convective vertical temperature difference is similar to curves first obtained in non-rotating models by SCHMIDT and MILVERTON (1935).

The lower branch, initially almost linear, curves and approaches some very small temperature difference (zero as far as these rough determinations can show) as an asymptote. At 30 watts input, this difference was about .02°C; at 50 watts input, it was less than .01°C. As far as these admittedly crude measurements can be interpreted, the curve shows no obvious discontinuities. This would seem to imply a continuous spectrum of eddies for thermal exchange extending from steady convection through increasingly unsteady convection to a regime of random, turbulent eddies.

APPLICATIONS TO METEOROLOGY

Cloud phenomena often considered to be examples of Benard cellular convection are commonly observed in the atmosphere (BRUNT, 1937). Altocumulus (especially the floccus variety) and cirrocumulus have been viewed as fields of Benard cells with ascending currents in their centers; cumulus mammatus and the rarer, net-like cirrus reticulatus are examples of cells with descending cores, the first a case in which the descending air remains saturated, and the second, descent in which dessication occurs. That one could say that it is the stabilizing influence of rotation which helps make these two atmospheric circulation types possible is quite unlikely since the changes in the required critical lapse rates in the atmosphere with rotation are extremely small (see below). The question of an explanation for these cloud phenomena remains open, but their analogues with respect to circulation direction have been shown in the laboratory and are available for further study. There are, however, several areas in which these ideas can be at least qualitatively applied to meteorological problems. As early as 1928 Jeffreys pointed out that rotational stability theory should be of importance to meteorology.

In connection with this discussion, the reader is reminded that the visible cloud usually constitutes only a small part of a convective cell, generally only a part of the cell's core. The clear spaces between individual clouds are the descending portions of cells, so that the average distance between clouds is probably a fair measure of the cell diameter. The height of the cell is the full depth of the convective layer, not merely the height of the visible cloud. For low level clouds of the cumulus variety, the convective layer usually extends from near the ground to some level above the height of the cloud tops, since the height to which visible clouds grow is more often limited by their entraining of drier air from the environment rather than by their reaching the upper limit of the unstable layer. In the examples treated below, we shall, for simplicity, assume that the cell core remains saturated and that the visible cloud does reach the top of the unstable layer. The cloud heights considered, then, are the maximum that could be attained provided that sufficient moisture was available.

It should also be noted that the rotation that plays the stabilizing role in the cumulus problem is the absolute rotation of the fluid in which the convection occurs. In a coordinate system on the face of the rotating earth, a rotation around the z axis analogous to the rotation of our cylindrical model is $\Omega_e \sin \phi$, the vertical component of the earth's rotation. As is often done, we specify that the motions occur in higher latitudes in order that the horizontal component be small. Thus we could write for $4\Omega^2$ in the Taylor number (22),

$$4\,(\Omega_e \sin \phi)^2 = f^2 \qquad (43)$$

where f is the Coriolis parameter. But more generally, the atmospheric fluid possesses a rotation of its own, relative to the earth; so a better approximation of the $4\Omega^2$ factor for atmospheric convection problems would be

$$(f+\zeta)^2 = Z^2 \qquad (44)$$

where ζ and Z are the relative and absolute vorticities, respectively. Since it does not imme-

diately follow that horizontal shear would lead to the same rotational effects, it is assumed that the major portion of the relative vorticity of the general current results from the curvature component. Thus we can write for the equations of motion

$$\frac{\partial u}{\partial t} - Zv = -\frac{1}{\varrho_0}\frac{\partial P}{\partial x} + \nu\nabla^2 u \qquad (45)$$

plus similar forms for the other two equations. In this way we substitute the local absolute vorticity for 2Ω in the Taylor number in order to apply the theory to atmospheric convection.

The Critical Lapse Rate

The concept of a critical Rayleigh number for convection has been extended to compressible fluids by JEFFREYS (1930). Assuming that the atmosphere is a perfect gas, we have the energy equation (Goldstein).

$$\varrho J c_p \frac{D\theta}{Dt} - \frac{Dp}{Dt} = \varrho J c_p \varkappa \nabla^2\theta + \Phi \qquad (46)$$

After perturbation analysis, the pressure term can be reduced to

$$-\frac{w}{\varrho J c_p}\frac{\partial p}{\partial z} = w\gamma_d \qquad (47)$$

where γ_d, the dry adiabatic lapse rate, arises by substitution of the hydrostatic relation. Writing $\partial\theta/\partial z = -\gamma$, we obtain the simplified perturbation energy equation

$$\frac{\partial\theta}{\partial t} - w(\gamma - \gamma_d) = \varkappa\nabla^2\theta \qquad (48)$$

where the primes have been omitted as before. Comparison with equation (12) shows $(\gamma - \gamma_d) = \beta$, a small amount which must be added to the dry adiabatic lapse before convection can begin. It follows that a static atmospheric layer with a lapse rate equal to the dry adiabatic is analogous to the isothermal liquid layer assumed in the theoretical analysis.

However, discussion of this slight increase over the dry adiabatic lapse rate that rotation makes necessary for the initiation of convection in the atmosphere becomes academic when the small size of β is realized. For a one kilometer deep layer of air with a mean temperature of zero degrees Centigrade, an absolute vorticity of 10^{-4} sec^{-1}, and eddy coefficients of viscosity and conductivity of the order of 10^3 cm^2 sec^{-1}, a calculation of R_c gives a required β of only 2.5×10^{-5} degrees per kilometer. If we ask for a set of conditions which will increase the lapse rate by just one per cent of the dry adiabatic, we obtain the improbable picture of a ten kilometer deep convective layer in which, if the eddy coefficients are of the order of 10^6, an absolute vorticity of 1.5×10^{-2} sec^{-1}, one hundred fifty times as large as the middle latitude Coriolis parameter, would be required.

Table 4

Absolute vorticity sec^{-1}	Eddy viscosity cm^2 sec^{-1}		
	10^3	10^4	10^5
10^{-4}	77 m	245 m	775 m
3×10^{-4}	45 m	141 m	447 m
10×10^{-4}	24 m	77 m	245 m

Minimum atmospheric depths required to increase the critical Rayleigh number by ten per cent for several values of absolute vorticity and eddy viscosity.

But appreciable percentage changes in the critical Rayleigh number can be caused by rotational influences of the order of those found in the atmosphere. Table 4 lists the minimum required depths for a ten per cent increase in R_c as indicated by the modified Stommel limit, $1.7(\nu/\Omega)^{1/2}$ considering several combinations of absolute vorticity and eddy viscosity. For minimum depths necessary to double R_c under these same conditions, multiply the tabular entries by 1.9.

Thus in atmospheric convection layers, the influence of rotation might be of some importance, if not in causing significant increases in lapse rates, then perhaps in bringing about secondary effects such as the increase of cells' depth-diameter ratios or the occurrence of overstable oscillations.

The Depth of Atmospheric Convection

The association of deep convection in the atmosphere with cyclonic relative vorticity has been noted by, among others, PETTERSSEN et al. (1946); they gave as a forecasting rule, "Deep and very deep convection (to heights above 600 and 400 millibars, respectively)

should not be forecast unless the upper air charts show indication of cyclonic (relative) vorticity". Of 73 cases of "deep convection", 62 were found to have occurred with cyclonic vorticity and only one with anticyclonic vorticity as measured at the 750 millibar level. In ten cases, the relative vorticity was either zero or indeterminate. Of 30 cases of "very deep convection", 28 occurred with cyclonic vorticity and none with definitely anticyclonic vorticity. Similar association was found by CRUTCHER et al. (1950) who, on a monthly, year round basis, found positive linear correlations from .73 to .89 between the cloud top height and the relative vorticity measured at 700 millibars in 454 observations on the Washington to Bermuda air route. This group pointed out that there was no adequate dynamic explanation for this association, especially in view of existing ideas which call for maximum positive vertical motion ahead of rather than on the trough line.

The expression for the depth-diameter ratio of convection cells as a function of the basic rotation (42) accurately predicts the measurements of ordinary cells in water. It can also be used for numerical examples which show a quantitative relation between deep atmospheric convection and areas of cyclonic relative vorticity.

For a shallow convection case, consider a D/d ratio of unity, applying for example, to fair weather cumulus in cells one kilometer deep. We shall use a reasonable and convenient value for the eddy viscosity, $10^{3\frac{1}{2}}$. From (42) we find $a_c^2 = 52.5$; equation (25) then determines $T = 4 \times 10^4$, and from (22) we determine the absolute vorticity of the general current $= 0.6 \times 10^{-4}$ sec^{-1}. In middle latitudes, the relative vorticity would be anticyclonic, -0.4×10^{-4} sec^{-1}.

For an illustration of deep convection, consider a D/d ratio of two, applying to cumulus congestus in cells five kilometers deep. A reasonable coefficient of eddy viscosity for the larger volume is $10^{4\frac{1}{2}}$. For these conditions, $a_c^2 = 212$, $T = 2 \times 10^6$, and the corresponding absolute vorticity of the general current is 1.8×10^{-4} sec^{-1}. The relative vorticity is cyclonic, 0.8×10^{-4} sec^{-1} in middle latitudes.

These are only fictional examples, designed for illustrative purposes; but it should be noted that reasonable cell dimensions and not unreasonable values of the eddy viscosity lead to vorticity values of reasonable order. It would be interesting to obtain actual D/d ratios for clouds and to test this relation quantitatively. The measurements required would be the average distance between clouds for the cell diameter, the height of the cloud tops above the ground as an approximation of the cell depth, and, of course, the absolute vorticity of the general current.

Clear and Cloudy Areas

An approximation of the Prandtl number for air in the case of atmospheric convection is the ratio of the eddy, rather than the molecular, coefficients of viscosity and conductivity. Even though the eddy values separately vary with the scale of the circulations involved, i.e., with the depth of the convective layer, their ratio is generally considered to retain an approximately constant value near unity. At any rate, it is of the same order as the critical Prandtl number, 0.677, but probably somewhat larger. This, combined with the fact that the required lapse rates (in excess of the dry adiabatic) for the two types of convection and the difference between them are very small gives rise to the possibility that both the ordinary convective and the overstable oscillating cellular modes might occur simultaneously in the atmosphere. With a large enough rate of heating from below, one sufficient to maintain the lapse rate required by R_c^* as well as that required by R_c, this would be possible.

Along with this possibility, we have the earlier observation that as the Taylor number increases, the critical wave number for the overstable mode gradually becomes smaller than that for ordinary convection. This effect is visibly evident in the relatively shifting minimums of the wave number curves in Fig. 7. So if the two convective types are both occurring at some non-zero value of the Taylor number, the wave number of the ordinary type will be somewhat greater than that of the oscillatory type and a "beat frequency" pattern of resonance and interference can develop.

Consider a field of simultaneously occurring convective types in a layer three kilometers deep under the influence of an absolute vorticity of 10^{-4} sec^{-1}, assuming an eddy viscosity of 10^4 cm^2 sec^{-1}. For these values we find that $T = 8 \times 10^5$, $a_c^2 = 153.$, and $a_c^{*2} = 94.4$. Assuming

that the two types occur in the form of superimposed fields of hexagons, we find for their respective depth-diameter ratios, $D/d = 1.71$, $(D/d^*) = 1.34$. Since they are both in a three kilometer layer, their average diameters are $d = 1.75$ km and $d^* = 2.24$ km. From these it can be determined that the central cores of the two types of cells closely coincide in widely separated areas located over the field at the vertices of equilateral triangles, 8.1 kilometers on a side. In the central areas of the triangles and along the middle sections of their sides, the two convective patterns interfere in varying degrees, the descending walls of either type damping the rising cores of the other. This model is consistent with the frequently observed distributions of convective clouds, e.g. trade cumulus, in relatively small cloudy areas with large clear areas in between. If the centers of the cloud groups of the eight kilometer triangular distribution are traveling ten meters per second, they will pass over or near a ground observer at eleven to thirteen minute intervals. This is an entirely reasonable period, according to observers who have timed the passage of groups of trade cumulus.

In areas of partial interference, the occurrence or non-occurrence of visible clouds would be determined by the resultant strength of the partially reinforced cores and the amount of moisture available for condensation. If conditions were such that only strongly reinforced convective cores could produce visible clouds, the pattern would be small cloudy areas separated by large clear spaces. Similarly, if only a small amount of convective lifting were needed, there would be large cloudy areas separated by small clear spaces. The tallest clouds would be found where complete resonance occurs, in the centers of the cloudy areas. Again, this model could be checked by aerial observation, measuring the distance between cloudy areas in addition to the distance between individual clouds, the depth of the convective layer, and the absolute vorticity of the general current.

The period of the overstable component of the interference pattern of the numerical example above, obtained by substituting in equations (37) and (38), is of the order of one hundred hours. Its large magnitude is due to the close proximity of the R_c and R_c^* curves in a fluid with a Prandtl number so close to the critical value. In other words, the nominally oscillating component would be practically steady throughout the life of the convection field. But the interference pattern would still be a slowly changing one since the other component, the ordinary cells, would be in an unsteady condition; for since the lapse rate must be large enough for R_c^*, and since the Prandtl number is greater than the critical value, it follows that the lapse rate would exceed the minimum requirements of R_c. One would expect an interference pattern undergoing random changes in time and space, much the way patches of altocumulus are seen to change.

Conclusion

In all of the foregoing illustrative examples, the only effects considered were those of the basic rotation of the fluid environment and the thermal instability produced by uniform heating from below. Obviously, the symmetric models that result could, at best, only rarely occur in the earth's notoriously non-uniform, non-stationary atmosphere. Horizontal variations in the rate of heating from below and the presence of horizontal and vertical wind shears are a few disturbing factors that are seldom absent.

In the model experiments, a deliberate effort was made to eliminate all influences other than those of the basic rotation and the vertical temperature gradient. The fluid was allowed to reach relative rest (solid rotation) before the convective circulations were begun, and uniform heat and cold sources at the bottom and top and careful insulation of the cylinder's sides kept the average horizontal temperature gradient near zero. With these precautions, the conditions of each of the rotating Benard cell experiments determined a point in the Rayleigh, Taylor number plane of Fig. 12.

In the "dishpan" experiments of Fultz, the Taylor number is still an appropriate rotational parameter, but the principal thermal gradient involved is horizontal. There is a horizontal analogue of the Rayleigh number for use with this model containing the horizontal instead of the vertical temperature gradient, a horizontal distance instead of the vertical depth, and a horizontal (centrifugal) acceleration instead of the vertical acceleration of gravity. The dishpan experimental conditions, then, can

be located as points on a "horizontal Rayleigh number", Taylor number plane.

These two planes bound a space in which the conditions for any rotating system with horizontal and vertical temperature gradients can be located. Fultz has pointed out that in such a space an appropriate coordinate combining the vertical and horizontal Rayleigh numbers could have the form of a Richardson number, the horizontal temperature gradients determining a vertical shear term through the thermal wind relations. Throughout the space, in analogy with the correspondence of specific areas of the vertical Rayleigh, Taylor number plane with regimes of heat transfer and modes of circulation, certain volumes would delimit the conditions for the occurrence of similarly specific phenomena. It was noted above that dishpan type wave motions occurred in the rotating Benard-cell experiments if the horizontal temperature gradients were not minimized. This is an example of an experiment whose characteristic point had moved out of the vertical Rayleigh-Taylor plane into a region of the three dimensional, vertical Rayleigh, horizontal Rayleigh, Taylor number space, a region that corresponded to wave motions. Similarly, dishpan experiment points appear to reach a volume of the space corresponding to Benard convection cells when the Taylor number is large enough, the horizontal temperature gradients are moderate, and the vertical temperature gradient is not zero.

Acknowledgements

The authors wish to express their sincere appreciation to Dr Dave Fultz and Dr George W. Platzman of the staff of the Department of Meteorology, The University of Chicago, for their continued assistance and encouragement. Similarly, they wish to thank Dr S. Chandrasekhar, of Yerkes Observatory, for his interest in the experiments and for several stimulating conversations with regard to the theoretical aspects of the study. Much of the original experimental apparatus and many subsequent modifications were due to Messrs. Walter Bohan and Robert Kaylor of the Hydrodynamics Laboratory.

REFERENCES

Benard, H., 1900: Les Tourbillons Cellulaires. *Rev. Gen. Sci. Pur. Appl.*, **12**, 1261, 1309.

— 1901: *Ann. Chim. & Phys.*, **23**, 62.

Brunt, D., 1951: Experimental Cloud Formation. *Compendium of Meteorology*, 1255—1262.

— 1937: Natural and artificial clouds. *Quar. J. Roy. Met. Soc.*, **63**, 277—288.

Chandrasekhar, S., 1953: The instability of a layer of fluid heated below and subject to Coriolis forces. *Proc. Roy. Soc.*, (A) **217**, 306—327.

Christopherson, D. G., 1940: Note on the vibration of membranes. *Quart. J. Math.*, **11**, 63—65.

Corn, J. and Fultz, D.: Synoptic analyses of convection in a rotating cylinder. Geo. Res. Papers, Geo. Res. Dir., A. F. Cambridge Res. Cntr.

Crutcher, Hunter, and Sanders, 1950: Forecasting the height of cumulus cloud tops. *Bull. Amer. Met. Soc.*, **31**, 1: 1—7.

Dorsey, N. E., 1940: Properties of Ordinary Water Substance. *Am. Chem. Soc. Mon.* **81**, Reinhold.

Eddington, A. S., 1926: *Internal Constitution of the Stars.* Cambridge U. Press.

Fultz, D., 1951: Experimental Analogies to Atmospheric Motion. *Compendium of Meteorology*, 1235—1248.

— and c.: Studies of thermal convection in a rotating cylinder with some implications for large-scale atmospheric motions and other problems. Hydrodynamics Laboratory, U. of Chicago (in preparation).

—, Nakagawa and Frenzen, 1954: Instance in thermal convection of Eddington's overstability. *Phys. Rev.*, **94**, 6: 1471—1472.

Goldstein, S., 1943: *Modern Developments in Fluid Dynamics.* 2nd edn., Oxford U. Press.

Jeffreys, H., 1926: The stability of a layer of fluid heated below. *Phil. Mag.*, **2**, 833—844.

—, 1928: Some cases of instability in fluid motion. *Proc. Roy. Soc.*, (A) **118**, 195—208.

—, 1930: The instability of a compressible fluid heated below. *Proc. Camb. Phil. Soc.*, **26**, 170—172.

von Karman and Biot, 1940: *Mathematical Methods in Engineering.* 1st edn., McGraw-Hill.

Pellew, A. and Southwell, R. V., 1940: On maintained convective motion in a fluid heated from below. *Proc. Roy. Soc.*, (A), **176**, 312—343.

Petterssen, Knighting, James, and Herlofson, 1946: Convection in theory and practice. *Geof. Pub.* **16**, 10.

Rayleigh, Lord, 1916: On convection currents in a horizontal layer of fluid. *Phil. Mag.*, **32**, 529—546.

Stommel, H., 1947: A summary of the theory of convection cells. *Ann. N. Y. Acad. Sci.*, **48**, 715—726.

Sutton, O. G., 1950: On the stability of a fluid heated from below. *Proc. Roy. Soc.*, (A) **204**, 297—309.

The Use of the Primitive Equations of Motion in Numerical Prediction

By J. Charney, The Institute for Advanced Study, Princeton[1]

(Manuscript received September 15, 1954)

Abstract

An obstacle to the use of the primitive hydrodynamical equations for numerical prediction is that the initial wind and pressure fields determined by conventional means give rise to spurious large-amplitude inertio-gravitational oscillations which obscure the meteorologically significant large-scale motions. It is shown how this difficulty may be overcome by the use of a relationship between wind and pressure which enables one to determine these fields in such a manner that the *noise* motions do not arise. The method is illustrated by a numerically computed example.

The wind-pressure relationship is in a sense a generalization of the geostrophic approximation and may be used where the latter approximation is inapplicable, either to determine initial conditions or to derive a set of *filtering* equations for numerical prediction analogous to the quasi-geostrophic equations.

I. Introduction

The recent widespread revival of interest in numerical weather prediction has been brought about partly by the development of high-speed electronic computing apparatus but also, to a large extent, by the very encouraging early results obtained through systematic use of the geostrophic approximation in what has come to be known as the quasi-geostrophic equations of motion. In these equations it is assumed that the wind is and remains nearly geostrophic, at least enough so that the horizontal acceleration terms in the equations of motion may be approximated by the accelerations of the geostrophic wind. Numerical integrations of these equations have already yielded forecasts comparable in accuracy to those produced by conventional methods, and probably superior in accuracy in the case of new developments. In purely theoretical investigations the use of the geostrophic approximation, by automatically filtering out high-frequency "noise" disturbances, has aided greatly in the understanding of the physical properties of large-scale flow systems where the high-frequency inertial effects are unimportant.

Yet those who have participated actively in this work are aware that this is barely the beginning of a new era, an era in which numerical methods will find increasing application in practical and theoretical meteorology, and that the quasi-geostrophic equations are but a first approximation to the equations which will eventually be used to describe the atmospheric motion. Experience has already

[1] This article is based on work done under contract N 6-ori Task Order I with the Office of Naval Research, U. S. Navy, and The Geophysics Research Directorate, Air Force Cambridge Research Center.

revealed systematic errors attributable to the geostrophic approximation. The approximation tends to displace depressions too far northward over the United States, and does not, apparently, permit the development of frontal discontinuities. It is of course possible to arrive at higher approximations by successive iteration of the geostrophic approximation, but this process, besides leading to severe mathematical complications, is valid only when the geostrophic deviations are small in the first place. Moreover, the time order of the equations is increased by one with each iteration, and since the time order cannot exceed *three*, it is evident that the sequence of approximations would not be convergent.

In any case it is likely that in the process the simplicity and elegance of the geostrophic approximation would be destroyed, and thereby also the reason for its introduction. One is therefore led to reexamine the possibility of using the primitive Eulerian equations, as Richardson[1] first proposed.

The difficulty here, as I have pointed out in previous papers, is that the initial values of wind and pressure cannot as a rule be prescribed independently with sufficient accuracy. One is apparently forced again to introduce for the wind values either the geostrophic approximation or something very little better. As will be shown later this not only introduces an initial inaccuracy but gives rise to spurious inertio-gravitational oscillations which obscure the meteorologically significant motions. The purpose of the present note is to show that accurate initial winds can be determined from the pressure field alone, and that if this is done inertio-gravitational oscillations will not arise.

II. Characterization of large-scale atmospheric motions

We shall deal with statically stable motions in which the horizontal scale far exceeds the vertical scale. It is therefore permissible to assume that the motions are in quasi-hydrostatic equilibrium. If f is the coriolis parameter $2\Omega \sin$ (latitude), V a characteristic horizontal velocity, and S a characteristic horizontal scale distance, the ratio of the inertial force to the coriolis force is in order of magnitude the non-dimensional Rossby number V/fS. The motion is said to be quasi-geostrophic if this number is small compared to unity. There are, however, cases where either because V is large or S is small the Rossby number is not small. Nevertheless the large-scale meteorologically significant motions are distinguishable from the inertio-gravitational oscillations, the other class of possible quasi-hydrostatic motions. We shall arrive at the required method for determining the horizontal wind field by asking for their distinguishing characteristics.

There are essentially two types of small amplitude wave perturbation of an atmosphere moving with constant angular velocity. They are distinguished by their frequencies and speeds of propagation. The meteorologically important type is characterized by small frequencies and by velocities of propagation of the same order of magnitude as the speed of the imbedding current. The second type has frequencies in excess of f, periods less than a half pendulum day, and velocities of propagation of the order of $(gH^2\partial \ln \theta/\partial z)^{1/2}$, where H is the vertical scale height of the disturbance and θ is the potential temperature. For the large scale disturbances of the atmosphere $H \sim 10^4$ m and $\partial \ln \theta/\partial z \sim 10^{-5}$ m^{-1}. Hence $(gH^2\partial \ln \theta/\partial z)^{1/2} \sim \sim \left(\frac{1}{10}gH\right)^{1/2} \sim 100$ m/sec^{-1}, which is much greater than the speed of the wind in the troposphere and very much greater than the speed of propagation of the observed large-scale systems. Even when the motions are of large amplitude and therefore not linearly superposable solutions of perturbation equations, a distinction can be drawn between the two types, at least in orders of magnitude of frequency and velocity of propagation.

From what can be inferred from available observations the bulk of the energy in the troposphere is confined to the first class of motion. This partition of energy may be attributed to the circumstances that the principle energy sources in the atmosphere have characteristic periods large compared to the pendulum day and therefore excite little of the second type of motion and that the first type is stable with respect to whatever perturbations of the second type are excited, either directly by the external sources of energy or indirectly by nonlinear interaction with the first type.

[1] Weather Prediction by Numerical Process, Cambridge, 1922.

III. The balance equations

Let us now examine the order of magnitude of the horizontal divergence in the meteorologically important systems—systems that are characterized by particle velocities and speeds of propagation small compared to the gravity-inertia speed $(gH^2\partial \ln\theta/\partial z)^{1/2}$. If a flow is quasi-geostrophic it can be shown by the use of dimensional arguments[1] that the ratio of the horizontal divergence $\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}$ to one of its component terms, say $\partial u/\partial x$, is just the Rossby number and therefore this is small. In the case where the Rossby number is not small the ratio is found to be $V^2/gH^2\partial \ln\theta/\partial z$ and is therefore again small. Thus the significant motions of the free atmosphere are characterized by small horizontal divergences, and we may therefore express u and v in terms of a stream function:

$$u = -\frac{\partial \psi}{\partial y},\quad v = \frac{\partial \psi}{\partial x}.$$

It may further be shown that the convective acceleration terms involving w in the horizontal equations of motion are smaller in order of magnitude than the remaining convective acceleration terms. If we omit these small terms, substitute the stream function expressions for u and v, and take the horizontal divergence of the equations of motion, we obtain, when pressure is used as the vertical coordinate,

$$\nabla^2\psi + \nabla\psi\cdot\nabla f - \frac{2}{f}\left[\left(\frac{\partial^2\psi}{\partial x\partial y}\right)^2 - \frac{\partial^2\psi}{\partial x^2}\frac{\partial^2\psi}{\partial y^2}\right] = \frac{\nabla^2\varphi}{f}$$

where φ is the geopotential. The above equation, which shall be called the *balance equation*, suffices to determine the horizontal velocity field if the field of geopotential is known. This equation is of the Monge-Ampere type and can be solved for ψ as a boundary value problem providing

$$\frac{\nabla^2\varphi}{f} + \frac{f}{2} > 0$$

i.e., if the geostrophic relative vorticity $\frac{\nabla^2\varphi}{f}$ is greater than $-f/2$. If the non-linear terms are small we obtain the geostrophic approximation by integration. In this sense the balance equation is a generalization of the geostrophic approximation. Both determine the velocity from the geopotential.

Now it is characteristic of the inertio-gravitational motion that the horizontal divergence is not relatively small. Hence if the u and v, which must be prescribed independently in the primitive quasi-static equations of motion, are determined from the balance equations, the inertio-gravitational motions are automatically excluded initially, and since they are not generated to any appreciable extent during the motion, they will never appear.

The balance equation has been derived by Fjørtoft (unpublished work) in an independent investigation as a necessary condition for the stability of the non-divergent flow with respect to perturbations with horizontal divergence. The neglect of the divergence terms in the divergence equation has been justified empirically by S. Pettersson[1] who finds them to be one to two orders of magnitude smaller than the remaining terms.

IV. Illustrations of the use of the balance equation

I had thought at one time that the use of geostrophic u's and v's as part of the initial specification of the flow would lead at most to small amplitude high-frequency inertio-gravitational fluctuations superposed on an essentially correct low-frequency trend motion. However a recent numerical integration of the non-linear Eulerian equations for a barotropic atmosphere carried out in collaboration with B. Bolin led to the suspicion that large amplitude inertio-gravitational perturbations can and do arise. In retrospect this is not surprising, for although the initial divergence of the geostrophic wind is zero or nearly zero, the first time derivative will not be zero since the balance equation will not be satisfied. The experiment was not conclusive because the fluctuations could have been due in part to computational instabilities arising from incorrect boundary conditions. (The boundaries were geometrical surfaces immersed in the fluid.) To investigate the matter further and to verify the filtering properties of the balance

[1] Charney, J. G., Geof. Publ. 17, No. 2, 1948.

[1] Tellus, 5, No. 3, 1953.

equation the following numerical experiment was recently performed in Princeton:

Initial conditions were prescribed for the flow of an atmosphere consisting of two homogeneous incompressible layers of different density bounded above and below by rigid surfaces. The lower layer was assumed to be so deep that its motion could be ignored. The upper layer would then move as though it were a single layer with a free surface but with gravity reduced in the ratio $\left(1-\frac{\varrho'}{\varrho}\right)$, ϱ' and ϱ being the densities of the upper and lower layers respectively. This was done to simulate the actual static stability of the atmosphere. The mean thickness of the upper layer was taken to be H and $\left(1-\frac{\varrho'}{\varrho}\right)$ was chosen to equal $H\partial \ln \theta/\partial z$ for the actual atmosphere, *i.e.*, about 10^{-1}. In this way the buoyancy forces, as well as the speed of the internal gravity waves, became comparable to those in the actual atmosphere. The coriolis parameter f was taken to be constant and was set equal to 10^{-4} sec^{-1} corresponding to a latitude of about 43°. The motion was assumed to take place in the x, y plane between the rigid lateral walls $y = 0$, $y = D$.

The initial velocity field was prescribed by the periodic stream function

$$\psi = A \sin\frac{2\pi x}{L} \sin\frac{\pi y}{D}$$

with L = 4,000 km and D = 2,000 km and $u_{max} = v_{max} = 2\pi A/L$ = 50 m sec^{-1}. To check first that the balance equation did indeed exclude inertio-gravitational motions the initial values of the thickness field h were determined from the balance equation, which in this case took the form

$$\nabla^2\psi - \frac{2}{f}\left[\left(\frac{\partial^2\psi}{\partial x\partial y}\right)^2 - \frac{\partial^2\psi}{\partial x^2}\frac{\partial^2\psi}{\partial y^2}\right] = \frac{\nabla^2\varphi}{f}$$

where $$\varphi = g\left(1-\frac{\varrho'}{\varrho}\right)h.$$

Since v, the y-component of velocity, is zero at $y = 0$ and $y = D$, the second equation of motion

$$\frac{dv}{dt} + fu = -\frac{\partial\varphi}{\partial y}$$

yielded the side condition

$$fu = -\frac{\partial\varphi}{\partial y}$$

at $y = 0$ and $y = D$. The solution of the balance equation subject to this condition was found to be

$$\varphi = g\left(1-\frac{\varrho'}{\varrho}\right)H + Af\sin\frac{2\pi x}{L}\sin\frac{\pi y}{D} + \frac{A^2}{4}\left(\frac{\pi^2}{D^2}\cos\frac{4\pi x}{L} + \frac{4\pi^2}{L^2}\cos\frac{2\pi y}{D}\right).$$

It is easy to show that if the divergence had vanished exactly the motion would have been exactly stationary. Since the use of the balance equation was supposed to prevent the occurrence of large divergences it was to be expected that also in the present case the motion would remain nearly stationary. This was found to be so for a forty-eight hour period to within the limits of a quite small computational error.

A numerical integration was next performed using the geostrophic initial conditions

$$\varphi = f\psi + g\left(1-\frac{\varrho'}{\varrho}\right)H,\quad u = -\frac{\partial\psi}{\partial y},\quad v = \frac{\partial\psi}{\partial x},$$

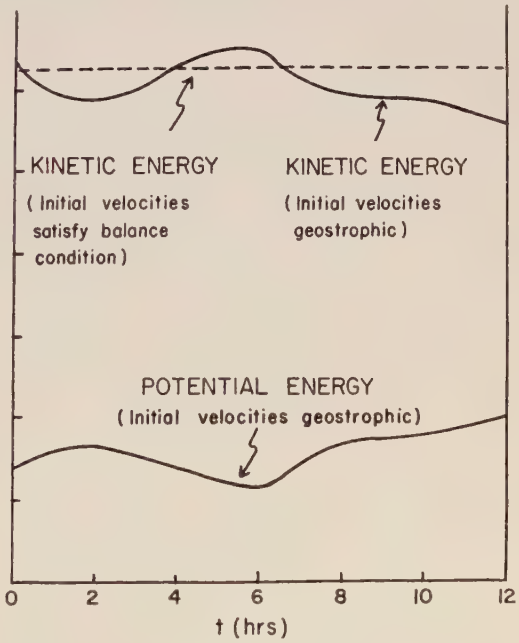


Fig. 1. Variations in time of kinetic and potential energy.

which of course also satisfy the boundary conditions $fu = -\partial\varphi/\partial y$ at $y = 0$ and $y = D$. In this case large divergences ocurred which in turn produced large potential energy changes according to the formula

$$\frac{dP}{dt} = \frac{\varrho}{2}\int_0^L\int_0^D g\left(1 - \frac{\varrho'}{\varrho}\right)\frac{h^2}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)dxdy$$

in which P is the potential energy per wave length. The kinetic energy was accordingly highly variable. Figure 1 shows the kinetic and potential energies plotted as functions of time. The upper curve represents the kinetic energy and the lower the potential energy. The horizontal line represents the constant kinetic energy in the non-divergent and balanced cases.

We note that there is a fluctuation of some 14 % in kinetic energy. The local fluctuations were considerably larger than this and must be regarded as errors.

The foregoing example indicates that the balance condition is far more suitable for determining the initial wind pressure fields than the geostrophic condition.

It should be remarked finally that the preservation of the velocity-pressure adjustment demanded by the balance equation necessarily entails horizontal divergences. I have merely attempted to show that these divergences must remain so small that the motion is approximately non-divergent at all times. Because the balance condition always holds, it is possible in principle to employ it in the same manner as the geostrophic approximation. It may be taken together with the hydrostatic equation, the vorticity equation, the physical equation and the continuity equation as determining the motion. The difficulty then is that it is not possible to combine these equations into a single equation in φ as was the case with the geostrophic approximation, and I have succeeded in solving them numerically only in the barotropic case (unpublished result).

Numerical Forecasting with the Barotropic Model[1]

By BERT BOLIN, University of Stockholm

(Manuscript received January 20, 1955)

Abstract

The experiments with numerical forecasting using the barotropic model have been continued and a number of 24, 48 and 72 hour forecasts are presented. The initial data for these forecasts covered an area of about 9,000 by 12,000 km. The results indicate that the boundary influences have been reduced to be unimportant in the centre of the area in the 24-hour forecasts but they may still cause errors in the 72-hour forecasts. In view of the approximations made very good 72-hour forecasts have been obtained in some cases. The most successful one gave a correlation of 0.87 and a relative error of the forecast height changes of 0.58. Most of these extended forecasts should be of definite value in forecasting the weather. Large errors are, however, still obtained in some cases. The neglection of baroclinic process is one of the reasons for these errors but it is quite obvious that the largest errors are of a different nature. Experiments are at present conducted to find the sources of these errors. One line of this research is a general re-analysis of the assumptions made in deriving the barotropic model which is presented in the latter part of this paper. A number of improvements are suggested for further tests.

1. Introduction

In a recent report (Staff Members, Institute of Meteorology, University of Stockholm, 1954, hereafter denoted by II) a number of 24-hour forecasts with the aid of the barotropic model were presented. In spite of the comparatively successful results it was concluded that several improvements in the computational procedure were desirable. In particular the assumptions on the boundaries of the forecast area affected the forecasts seriously in some cases. It was not possible to use a larger area than about $5{,}700 \times 5{,}700$ km with a gridsize of 300 km, because of the limited capacity of the Swedish Computer BESK at that time. Nor was it possible to extend the forecasts beyond 24 hours for the same reason. Since then a magnetic drum capable of holding 4,096 40 binary digit words has been added to the machine, which has made it possible to extend the forecasts to cover an area more than three times as large as the one previously used and also accordingly increase the forecast period. A number of such forecasts have now been made and the results of those are presented here.

The first series of such forecasts over two and three days immediately made it clear that it was highly desirable to make such tests over a longer period both from the point of view of using them in daily weather forecasting routine as well as for the analysis of different kinds of errors. Such an extensive testing program therefore was started by the Weather Service of the Swedish Air Force. Since December 1, 1954 more or less regular 24, 48 and 72-hour forecasts are made. As a com-

[1] The research reported in this document has been sponsored in part by the Geophysics Research Directorate of the Air Force Combridge Research Center, *Air Research and Development Command, United States Air Force*, under contract No. AF 61 (514)-648-C, through the European Office, ARDC.

parison the Swedish Weather Bureau prepares forecasts with conventional extrapolation methods. Some of the preliminary results of these forecasts are considered in the discussion here.

By and large the results are very encouraging and even the forecasts over 72 hours ought to be useful in actual weather forecasting in most cases. In the centre of the forecast area the effects of the assumptions on the boundaries seem to be small. However, many systematic errors still appear as can be seen from the presentation below. In a few cases these errors are obviously due to the fact that we disregard baroclinic effects completely, but in most cases there seem to be other reasons for the disagreement between the computed and the real development. Among other things it therefore seems very desirable to reconsider the approximations within the barotropic model and analyze the effects of them. Such an analysis is attempted in the latter part of this paper.

Fjørtoft (1952) has developed a graphical method for solving the barotropic vorticity equation. Since this method is very simple and can be used in the daily forecasting routine in any weather service without having an electronic computer at disposal it is of very definite interest to compare the results obtained in such a way with the results of computations on a machine. Such a comparative study has been made by O. Haug and is also presented here.

The initiation of this work has been possible by grants from the Knut and Alice Wallenberg Foundation and the Swedish National Science Research Council and by the support from the Office of Naval Research, USA. The continuation has been supported by the Weather Service of the Swedish Air Force and the Air Research and Development Command of the United States Air Force. The practical work of carrying through the forecasting program discussed here was done by a team consisting of H. Bedient, USA; P. Bergthorsson, Iceland; G. Dahlquist, Sweden; B. Döös, Sweden; O. Haug, Norway, A. Jonasson, Sweden, and the author. A number of discussions with Drs. P. Welander and G. Dahlquist have contributed to the general discussion in the latter part of the paper. All these contributions are gratefully acknowledged. Finally I am greatly indebted to Professor C.-G. Rossby for his neverfailing interest in this work and for many stimulating discussions.

2. Methods of computations and truncation errors

In the attempts to use the barotropic model over a larger area and over periods of two or three days about the same procedure at first has been used as given by Charney and Phillips (1953) and which also was used in II. A few minor changes have been made:

The criterion for computational stability puts an upper limit on the length of the time-steps, which is about one hour if the gridsize is 300 km. The reason for this is that only the neighbouring gridpoints are used when computing the change in time at a point with the aid of finite differences in space. It is quite obvious that the time derivative thus obtained only can be representative for a time period which is equal to or less than the time it takes for influences to travel the distance between these neighbouring grid points and the point itself. To be able to take longer time steps it is necessary to use a method by which influences from points at larger distances automatically are included if necessary. The following method has been tried: In order to compute the change of vorticity at a certain point, (x_0, y_0) from the time t to $t+\Delta t$, we try to find the location (x', y') of a particle at time t, which at time $t+\Delta t$ will be located at (x_0, y_0). The difference between the vorticities in the two points (x', y') and (x_0, y_0) is then equal to the vorticity change in (x_0, y_0) over the time interval Δt. The size of Δt is here limited by the accuracy with which the coordinates (x', y') can be determined. By using second differences in the interpolation schemes it has been possible to use time steps of 3 hours in a field where the maximum velocity of the flow was 60 m/sec. It was believed that this increase of the time step would shorten the total time for a forecast but the computational method itself is considerably more complicated than a Jacobian computation. Furthermore the points close to the boundary require a special handling which still more increases the length of the computations. No gain in time was thus obtained with this method. Instead the following scheme was used in the final computations (devised by G. Dahlquist).

The basis of the method for the integration of the equations

$$\frac{\partial \xi}{\partial t} = \frac{\partial \phi}{\partial y} \cdot \frac{\partial \eta}{\partial x} - \frac{\partial \phi}{\partial x} \cdot \frac{\partial \eta}{\partial y} = J(\eta, \phi) \qquad (2.1)$$

$$\eta = \frac{m^2}{f} \xi + f \qquad (2.2)$$

$$\xi = \nabla^2 \phi \qquad (2.3)$$

is the trapezoidal rule

$$\xi^\tau - \xi^{\tau-1} = \Delta t \cdot J\left(\frac{\eta^\tau + \eta^{\tau-1}}{2},\ \phi^{\tau-1/2}\right) \qquad (2.4)$$

and the usual finite difference approximation

$$\xi^\tau = (\bar{\phi}^\tau - \phi^\tau) \cdot \frac{4}{(\Delta s)^2} \qquad (2.5)$$

Here ξ^τ etc. denotes the value of ξ at the time $\tau \cdot \Delta t$, the bar denotes the average of ϕ^τ in the four neighbouring lattice-points and Δs is the mesh-width (which is equal to 300 km at 50° latitude). Other notations as in (II). Since ξ^τ occurs implicitly on the right hand side of (2.4) the following modified scheme has actually been used.

The height values ϕ° are given initially. Then ξ° is computed from (2.5) and η° is obtained from (2.2).

We get

$$\xi_a^{1/2} = \xi^\circ + \frac{1}{2} \Delta t \cdot J(\eta^\circ, \phi^\circ)$$

where the subscript a denotes that the value is considered as a first approximation to the quantity in question. Then $\eta_a^{1/2}$ is computed from (2.2), and $\phi^{1/2}$ is obtained from $\xi_a^{1/2}$ through the solution of Poisson's difference equation (2.3). Here we utilize Liebmann's method of iteration taking ϕ° as the first approximation.

After this, the computation of the field at $\tau = 1, 2, 3, \ldots$ can proceed step by step in the following manner. Compute

$$\xi_a^\tau = \xi^{\tau-1} + \Delta t \cdot J(\eta_a^{\tau-1/2},\ \phi^{\tau-1/2})$$

$$\xi^{\tau-1/2} = \frac{1}{2}(\xi_a^\tau + \xi^{\tau-1})$$

Then $\eta^{\tau-1/2}$ is obtained from (2.2), and the final value of ξ^τ is computed from

$$\xi^\tau = \xi^{\tau-1} + \Delta t \cdot J(\eta^{\tau-1/2},\ \phi^{\tau-1/2})$$

For the use in the next step we also extrapolate

$$\xi_a^{\tau+1/2} = 2\xi^\tau - \xi^{\tau-1/2}$$

and compute the corresponding $\eta_a^{\tau+1/2}$. Then ϕ^τ is solved from (2.3), using the linear extrapolation $2\phi^{\tau-1/2} - \phi^{\tau-1}$ as a first approximation. Finally we extrapolate

$$\phi^{\tau+1/2} = \phi^\tau + \frac{1}{2}(\phi^\tau - \phi^{\tau-1})$$

and are ready for the next time step.

In the Liebmann process for solving (2.3) an over-relaxation factor of 1.2 was used. The solution was accepted when $| 1/4\ (\Delta s)^2\ \xi^\tau - (\bar{\phi}^\tau - \phi^\tau) |$ nowhere exceeded 0.3 m. As an average four or five iterations were needed. It can be shown that the Courant-Friedrich-Lewy criterion for computational stability is changed somewhat for a procedure as outlined above and the largest permissble value of Δt increases with 40—50 %. In most of the forecasts made here a time step of 1 ½ hour was used. The total time for a 24 hour forecast using 1,240 grid points was around 15 minutes.

It is quite obvious from these computations that 72 hours or somewhat more represents the limit with the method used. In most cases quite large oscillations in the vorticity field appear at this time, as can also be noticed in the height field. This is probably not due to computational instability in the sense of Courant-Friedrich-Lewy. A few forecasts were made using different time steps and no appreciable difference was obtained if the modified stability criterion mentioned above was satisfied. Instead these small scale disturbances are probably gradually activated through nonlinear interactions, which is a normal feature of any such nonlinear process. The behaviour of such short waves is very badly represented by the finite difference approximations used. It therefore seems necessary to introduce systematic smoothing of the height field (or vorticity field) when extending the forecasts beyond about 48 hours. Such a procedure will be tried during the testing now in progess.

3. Forecasting with the barotropic model over 24, 48 and 72 hours

The results of the forecasts are summarized in table 1, r is the correlation coefficient between observed and computed changes, σ_x and σ_y are the root mean squares of the observed and computed changes, and ε is the mean error. The verification as presented here

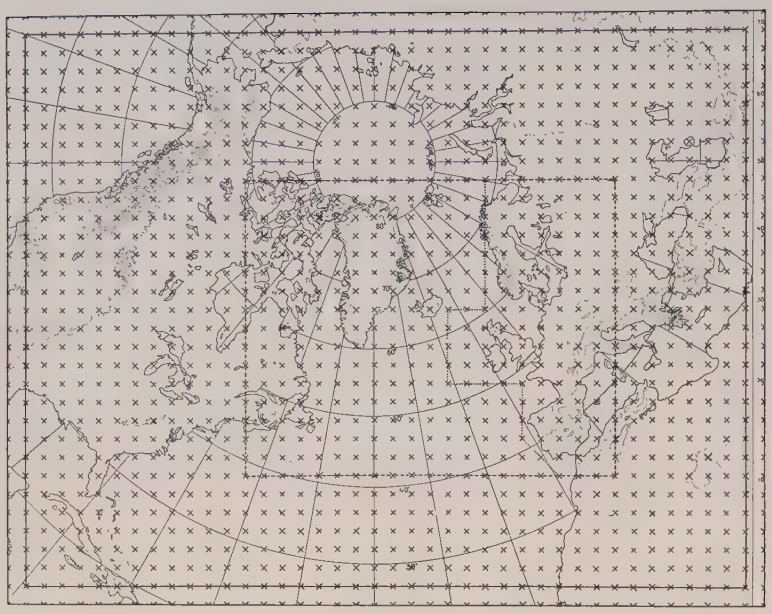


Fig. 1. Location of the grid used in the forecasts. The computations were made over the area limited by the inner solid line. At the initial time data was supplied also at the points outside to get an accurate value of the vorticity on the boundary. The verification was done over the area inside the dashed line and also over the land area indicated by the dotted line in the eastern part of this area.

was made over a comparatively small area over western Europe as indicated by the dotted line in fig. 1. The corresponding values have also been computed for a larger area indicated by the dashed lines in fig. 1. The average values of this latter comparison are given at the bottom of the table. The statistical quantities given in table 1 do not give a complete picture of the success or failure of the forecasts, but will still be useful for this discussion. To judge a forecast completely a careful inspection of the change maps in relation to the flow pattern is necessary. One such example will be given in the next section.

Table 1

Date	24^h barotropic forecast					48^h barotropic forecast					72^h barotropic forecast					48^h forecast; conventional methods				
	r	σ_x	σ_y	ε	ε/σ_x	r	σ_x	σ_y	ε	ε/σ_x	r	σ_x	σ_y	ε	ε/σ_x	r	σ_x	σ_y	ε	ε/σ_x
24/11-51 03	0.86	113	94	60	0.53	0.91	187	174	78	0.42	—	—	—	—	—	—	—	—	—	—
25/11-51 15	0.94	131	123	46	0.35	0.90	174	186	78	0.45	0.68	115	161	127	1.10	—	—	—	—	—
27/11-51 03	0.87	117	124	60	0.52	0.77	140	183	109	0.78	—	—	—	—	—	—	—	—	—	—
1/1 -54 03	0.78	68	60	53	0.79	0.63	121	118	105	0.87	—	—	—	—	—	—	—	—	—	—
21/9 -54 03	0.86	65	47	34	0.52	0.84	89	83	48	0.58	0.63	90	95	82	0.91	—	—	—	—	—
25/9 -54 03	0.91	87	81	40	0.46	0.84	110	121	84	0.76	—	—	—	—	—	—	—	—	—	—
26/9 -54 03	0.81	61	57	36	0.60	0.78	78	93	67	0.86	0.71	93	128	96	1.03	—	—	—	—	—
27/9 -54 03	0.60	46	37	40	0.85	0.74	81	64	58	0.72	0.63	89	92	87	0.98	0.70	81	56	84	1.04
28/9 -54 03	0.76	67	67	47	0.70	0.74	81	83	60	0.74	0.58	76	49	62	0.82	0.74	81	58	59	0.73
29/9 -54 03	0.93	71	71	28	0.39	0.84	94	85	55	0.59	0.62	136	115	113	0.83	0.80	94	46	65	0.69
30/9 -54 03	0.89	78	65	37	0.48	0.86	133	103	66	0.49	0.72	169	142	114	0.67	0.57	133	147	125	0·94
1/10-54 03	0.92	91	55	40	0.51	0.88	140	129	65	0.46	0.85	167	148	81	0.49	0.77	140	116	124	0.89
2/10-54 03	0.92	81	87	33	0.41	0.92	143	121	62	0.43	0.87	146	145	85	0.58	0.83	143	175	98	0.68
Mean	0.85	83	74	43	0.52	0.82	121	119	72	0.59	0.70	120	119	95	0.79	0.74	112	100	92	0.82
Verification over larger area	0.75	85	74	56	0.66	0.71	116	119	91	0.78	0.58	121	134	123	1.02					

The following points should be mentioned specifically:

1) On the whole the agreement between observed and computed changes was worse over the larger area than the one only covering parts of the European continent and adjacent ocean areas. In the western parts the influences from the boundaries may have been of some importance after 48 and 72 hours, but already in the 24-hour forecasts a significant difference shows up. Undoubtedly this is to some extent the result of the comparatively few radiosonde observations over the ocean. However, another effect is also important. Computing the average errors in the 48-hour forecast of the height changes we find that they had a maximum in the vicinity of New Foundland and Nova Scotia (average of 13 forecasts). The computed changes were here on an average 80—100 m too high. This means that the absolute vorticity in reality was larger in this region than was computed. Since the barotropic model assumes conservation of the absolute vorticity this result indicates that this region in reality is a source region of vorticity. This result is not surprising. From experience we know that most cyclones are formed along the American sea border where the baroclinicity of the atmosphere on an average is large. Furthermore the thermal contrast between the continents and the ocean is larger at the American east coast than at the European west coast. *The results obtained indicate that baroclinic processes and non-adiabatic influences from the surface of the earth are less important over the eastern Atlantic and western Europe than over the American east coast and western Atlantic for forecasting changes of the 500 mb flow.* Of course this is not necessarily true in individual weather situations, but only expresses an average condition.

2) The 48-hour forecasts compare well with the 24-hour forecasts and also those for 72 hours were in most cases of definite value. It is interesting to notice that in some cases the correlation coefficient was larger and the relative error ε/σ_x smaller for the 48 hour forecasts than for those over 24 hours. On September 27 this was partly a result of a deepening cyclone over the Baltic during the first 24-hour period, which moved out of the area during the following day. However, on November 24, 1951 it was not the result of such a development. Instead we see that σ_x is considerably greater for 48 hours than for 24 hours (as in all cases presented here). This means that the quasi-periodic changes caused by the large scale patterns in the atmosphere on an average have a half period which is larger than one day. In the changes over 24 hours the large scale changes thus play a relatively less important role than those with shorter period since they have not yet reached their maximum intensity. However, the barotropic model is less applicable to systems of smaller scale and furthermore the finite difference approximations become worse, which explains this initial increase of the correlation coefficient. A series of 12-hour forecasts would probably show consistently lower correlation coefficients than those here obtained for 24 hours.

3) It is to be expected that the small scale atmospheric systems are poorly described by the barotropic model and the finite difference method also introduces large errors in computing the changes of such systems. This is clearly seen from an inspection of the vorticity field. The amplitude of the vorticity is considerably larger in she small systems than in the larger ones, while still the latter are most important for determining the general flow. Even in the very best forecasts there are large differences between the observed and forecast vorticity fields. This fact speaks in favour of the introduction of a systematic smoothing in the course of the computations.

4) The notion that the barotropic model describes large scale process in the atmosphere better than those of smaller scale seems not to be valid for the very largest systems. This is indicated by the fact that the correlation between computed and observed *mean* changes over the verification area in these 13 forecasts is somewhat smaller than the correlation of the changes at different points within the area.

5) The results presented here for 24 hours are on an average better than those given in the previous report II. Four synoptic situations have been treated both with the smaller area of initial data used in II and the larger one used here. For the three November cases in 1951 the differences in the results are hardly significant. Those made with the larger area seem to be slightly better. The forecast from January 1, 1954 was previously a failure with a correlation of 0.17 and a mean error of 169 m. It

was quite obvious that the boundary assumptions to a large extent were responsible for this result. The computations using the larger area resulted in an improvement to the values $r = 0.78$, $\varepsilon = 53$ m. The verification was not done over quite the same area but the values are still comparable.

6) A study of the error maps shows that this field to some extent is associated with the existing flow pattern. A knowledge of the error field over one 24-hour period permits us to draw some conclusions concerning the errors during the following 24-hour period. In whatever way these errors arise it should be possible to utilize the information they contain for improving the following forecast.

7) The last eight forecasts (September 25—October 2) were made in cooperation with the Weather Service of the Swedish Air Force and used in their daily routine forecasting. During the same period 48-hour forecasts were made by the Swedish Weather Bureau with conventional extrapolation methods similar to those developed by SCHERHAG (1948). The results of those are also given in table 1. It should be remarked that no major baroclinic development occurred during this period which at least to some extent might have been caught by the conventional methods, but certainly not by the barotropic model. It is of course also conceivable that some other method of extrapolation may give better results than the one used here. For 24-hours the conventional methods seem to have given somewhat better results than the barotropic model (not shown in the table). However, for 48 hours the barotropic model definitely was superior in these cases.

8) Very large errors sometimes appear after about 48 hours. They are among other things characterized by the fact that the subtropic anticyclones are intensifying considerably as well as the polar cyclones resulting in a general intensification of the gradients. In a few cases absurd values are obtained for the geopotential field as for example the value of 6,160 m in the centre of the anticyclone over the Eastern United States in the 72 hour forecast from October 2 (fig. 5). In the same forecast one also finds quite a strong rise over most part of the large verification area towards the end of the forecast period. Thus the computed mean rise over this area from 24 to 72 hours was more than 100 m, a value never observed in reality over such large an area.

It is quite clear that these large scale errors hardly can be the result of baroclinic developments in the real atmosphere. On the other hand the non-barotropic processes associated with the influences from the surface of the earth seem to be of some importance in forecasts over 48 hours or more. However, correcting for this in a similar way as proposed by CLAPP (1953) still leaves large errors unexplained.

Again we come back to the boundary assumptions. Keeping the vorticity at a fixed value on the boundary may cause large errors to build up gradually. Let us for example assume that a comparatively low value of the absolute vorticity is assigned to a portion of a southern boundary across which inflow takes place. This value is then kept fixed in the course of the following computation. The flow north across the boundary turns anticyclonically to the right (northern hemisphere). Here gradually an anticyclone will form meaning an intensification of the southerly flow to the west. Thus, a day later the vorticity carried in across the boundary is brought further to the north than on the previous day and also the anticyclone extends further to the north. At these higher latitudes the same small absolute vorticity coming from the boundary means a still more intense anticyclonic circulation because of the variation of the Coriolis parameter. In view of such effects *it may in many cases be necessary to place the boundaries in such a way that no flow takes place across them, if the forecasts are to be extended beyond 48 hours or else modify the assumptions on the boundaries to exclude developments of this kind.*

It still does not seem likely that all major errors in the forecasts will be removed by corrections of this type. We notice for example in the 48 hour forecast from October 2, 1954 (fig. 2) an anticyclogenesis over the sea between Greenland and northern Scandinavia, which did not occur in reality. This may be the result of the fact that the cyclonic vorticity in the cyclone to the east of the British Isles was over-estimated by the geostrophic approximation, which caused a turning of the flow more sharply than actually occurred for example after 48 hours (fig. 4). This error in the evaluation of the vorticity essentially

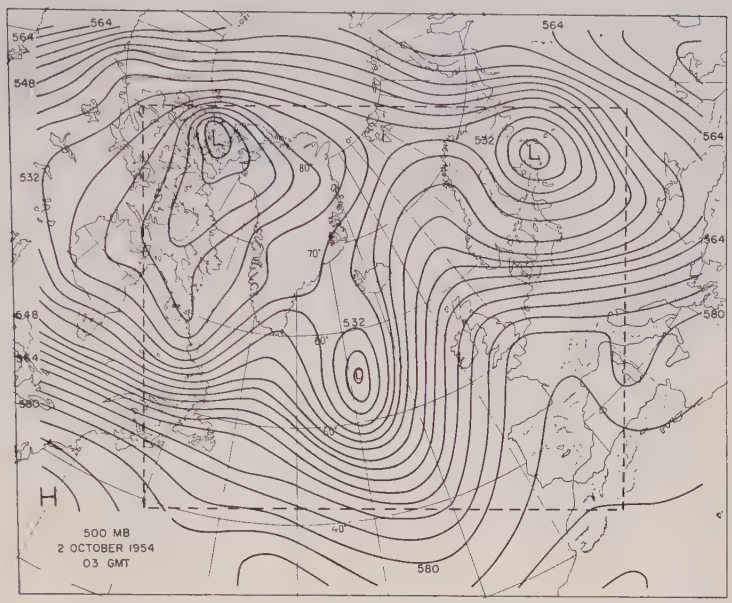


Fig. 2. 500 mb contours on October 2, 0300 GMT 1954. The heights are given in decameters.

depends upon the curvature of the flow and is small in the case of straight flow. Therefore, with the assumption of conservation of the absolute vorticity the *geostrophic* vorticity is not conserved for a particle at one time being in a region of strong curvature at another time in more or less straight flow.

The series of questions raised above makes it clear that a reanalysis of the barotropic model is highly desirable. We shall return to this problem in section 5.

Forecasts with Fjørtofts graphical method

by O. Haug

For the period 25 September to 2 October 48 hour forecasts have been made by means of Fjørtofts method in order to compare the results thus obtained with those obtained numerically with the aid of BESK. For details of the method we refer to the paper by Fjørtoft (1952). The computations were here done using a modified formula, which also takes into consideration the changes in the mean field, while the mean of the mean field is kept fixed in time. This method will be discussed by Fjørtoft in a forth-coming paper. The gridsize used for constructing the mean map was 600 km at 50° N and for constructing the mean of the mean 1,200 km.

The prognoses from the first four days were about as successful as those made with the machine. The correlation coefficients for the observed and computed changes evaluated over the large area indicated in fig. 1 are given in the following table.

Table 2

	25—27 sept.	26—28 sept.	27—29 sept.	28—30 sept.
BESK	0.58	0.77	0.64	0.83
Fj. method ..	0.70	0.60	0.60	0.71

During the later period there was a very pronounced increase of the amplitude of the large scale wave pattern over the Eastern Atlantic and Western Europe. The forecasts by BESK became better (see table 1) and apparently the development was to a considerable extent barotropic. The forecasts with the graphical method, on the other hand, were not as good, which in particular was true regarding the forecast of the amplitude of the waves.

The main source of errors in the forecasts by Fjørtoft's method seem to be the actual variations of the mean of the mean field, which is supposed to be constant during the forecast interval of 48 hours. During the first part of the period only small changes in the mean of

the mean field occurred, while it changed considerably during the later part. There are, however, reasons to believe that these pronounced changes were quite extreme. Such changes do not occur very often, but they usually represent major changes of the weather in the area concerned and are extremely important for the weather forecasting.

The displacement of troughs and ridges were in most cases fairly well forecast by Fjørtoft's method. The error in displacement of the trough and ridge, moving from the Atlantic in towards Europe during the later part of the period are given in the following table 3 as the percentage of the time displacement. The figures represent the mean of the errors of displacement at the latitude circles 60° and 50° N.

Table 3

Percentage error in prognosis of displacement of trough.

	29 sept. −1 oct.	30 sept. −2 oct.	1−3 oct.	2−4 oct.
BESK	9 %	20 %	7 %	26 %
Fj. method ..	23 %	77 %	18 %	33 %
Percentage error in prognosis of displacement of ridge				
BESK	27 %	21 %	7 %	10 %
Fj. method ..	10 %	18 %	17 %	4 %

It may be of some interest to see that the trough displacement was considerably better forecast by BESK, while the motion of the ridge was very well forecast using graphical methods, even better than with the machine.

It should finally be added that it takes 4—5 hours for one man to prepare a 48-hours forecast with Fjørtoft's method. The time can be reduced considerably if two persons work simultaneously.

4. Barotropic forecasts from October 2, 1954, 03 GMT

A correlation coefficient or mean error certainly gives a very poor description of the success or failure of a forecast. We shall therefore in some detail discuss some of the forecasts made in this series and have chosen those based on the map of October 2, 1954, 03 GMT. It was the most successful forecast over northwestern Europe for 48 and 72 hours and it is interesting in view of the prevailing synoptic situation. On the other hand the results for the western parts of the verification area (western Atlantic) were worse than on the average (fig. 2—5).

The initial map was characterized by two well defined troughs, one along 30° E, the other one at 25—30° W. The ridge between these troughs had been intensifying during the last 24 hours. Further west another trough was situated on the lee side of the Rocky Mountains. Off the American west coast a very intense blocking ridge had remained stationary for some days (not shown on the map). During the following 48 hours the ridge over the Bristish Isles intensified and moved east and the trough further west also advanced quite rapidly eastwards. As is seen from fig. 4 the 48 hour forecast of the positions of the trough and ridge was almost perfect. The very characteristic changes in the shape of the ridge in that the distance between the two troughs on both sides became smaller is clearly indicated in the forecast. The westerly current from the Atlantic pushing westward towards central Europe had reached the zero meridian on October 4 in the morning. In the forecast it had advanced somewhat further and a westerly current was gradually being generated over central Europe. During the following 24-hour period this development continued and the weather was radically changed in that humid maritime air replaced the dry polar air that had earlier been brought in from the north over most parts of central Europe. It is very interesting indeed that such a development of the decline of the ridge and the intensification of the anticyclone over northern Finland and Russia in principle could be forecast three days in advance. It is also interesting to notice that the correlation between the observed changes between October 4 and 5 and those computed for the same 24-hour period using the initial map two days earlier was 0.65.

In other regions the errors in the forecasts were larger. Over the Norwegian Sea the current never turned back towards NNW as was forecast. As mentioned in the previous section this possibly may be explained as a result of an overestimation of the cyclonic vorticity in the trough at 30° W on October 2. The depth of the centre may also have been

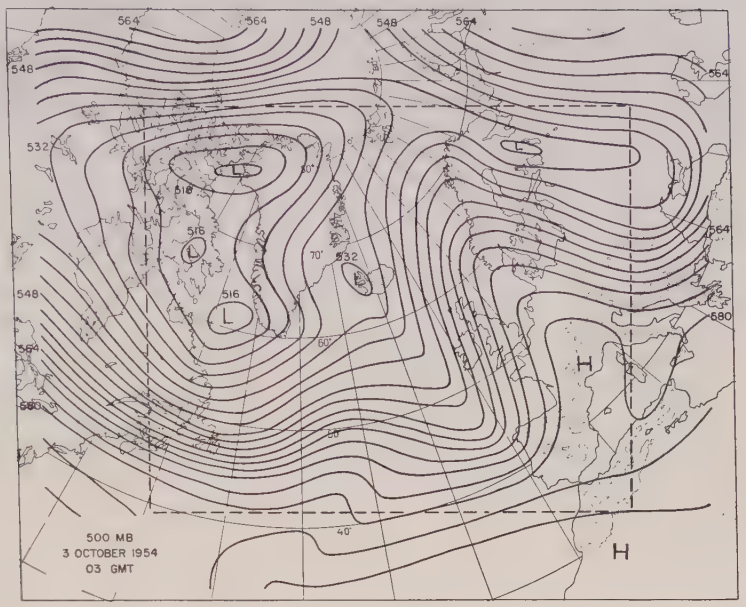


Fig. 3 a. 500 mb contours on October 3, 0300 GMT, 1954.

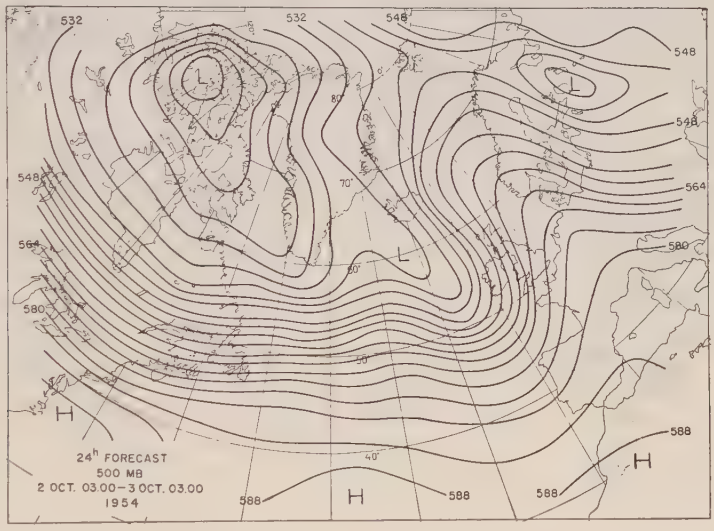


Fig. 3 b. 24-hour forecast of 500 mb contour field from October 2, 0300 GMT to October 3, 0300, 1954.

erroneously analyzed as no observations existed close to it.

To get an idea of the importance of the geographically fixed errors the following computation was carried through. The 48-hour forecast was corrected with the aid of the mean 48-hour errors found from the other 12 forecasts given in table 1. In this way the correlation coefficient over the smaller verification area was changed from 0.92 to 0.93 and the mean error from 62 to 59 m. The improvement is hardly significant. Probably a more pronounced improvement in general would be obtained at the east coast of America. Furthermore it would of course be better to introduce corrections of this type in the course

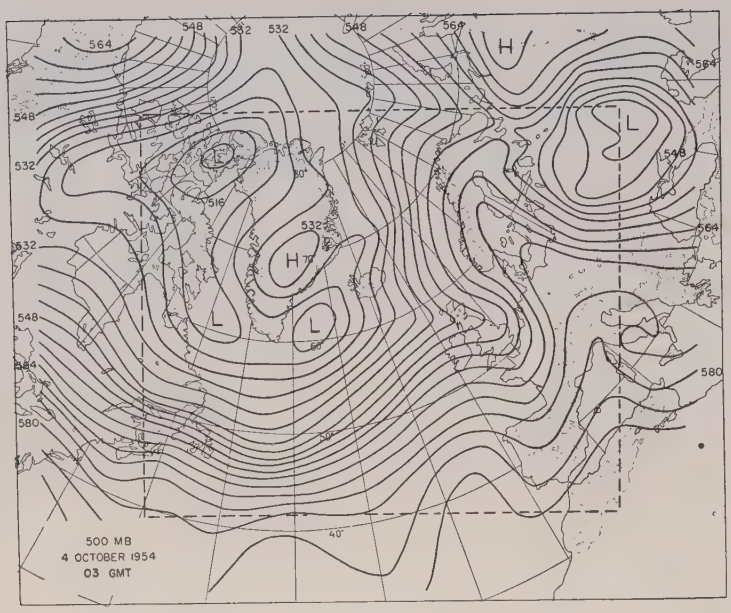


Fig. 4 a. 500 mb contours on October 4, 0300 GMT, 1954.

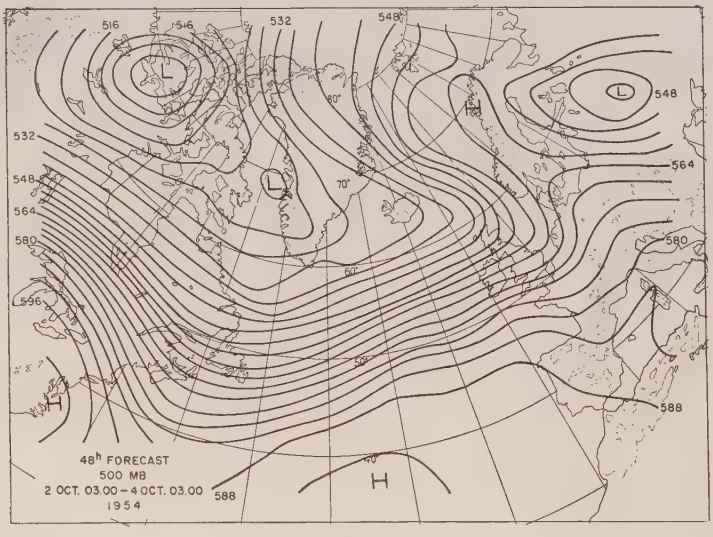


Fig. 4 b. 48-hour forecast of 500 mb contour field from October 2, 0300 GMT to October 4, 0300 GMT, 1954.

of the computations, since such errors cause new errors if not continuously removed.

5. The barotropic model

There are a number of reasons, why the barotropic, two-dimensional model was chosen for the first attempts in numerical forecasting. The quasi-two-dimensional character of the motion of the atmosphere had been stressed repeatedly by Rossby and his collaborators (e. g. Rossby 1939). The transformation of potential energy into kinetic energy only amounts to some 10—20 % of the total kinetic energy of the atmosphere per day and to a first approximation the changes of atmospheric flow-pattern represent a redistribution of

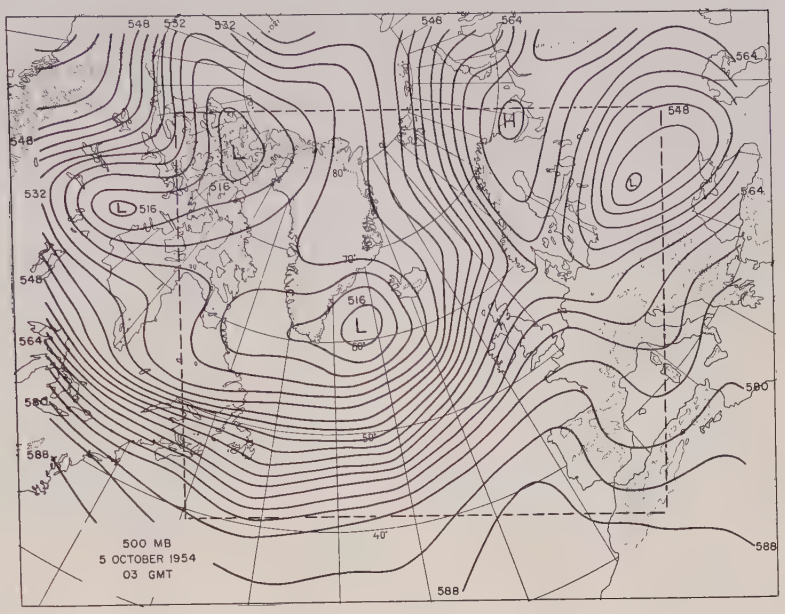


Fig. 5 a. 500 mb contours on October 5, 0300 GMT, 1954.

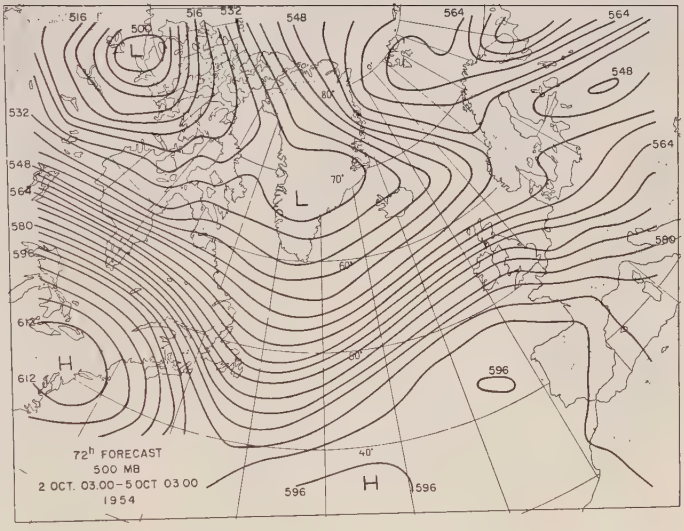


Fig. 5 b. 72-hour forecast of 500 mb contour field from October 2, 0300 GMT to October 5, 0300 GMT, 1954.

kinetic energy. Above all it is important to start from the simplest possible idea about the dynamics of the atmosphere and gradually proceed to more complicated models. In doing so we can get a better understanding of the relative importance of various processes in the atmosphere. In that sense the barotropic model offers an excellent starting point.

There are several advantages in using an internally consistent model of the atmosphere in such an approach. It is easier to visualize the behaviour of the field of motion and to clarify the importance of various *physical* factors. In starting from the general equations in three dimensions and attempting to derive a set of forecast-equations by successive ap-

proximations inconsistencies are easily introduced, the importance of which is difficult to determine.

For this reason one has chosen a homogeneous and incompressible fluid of a finite depth. We shall study it both assuming a fixed upper surface of this fluid (i.e. put div $\mathbf{v} = 0$) as well as a free surface. The following three equations describe the motion of the fluid with a free surface.

$$\left.\begin{aligned} \frac{du}{dt} - fv &= -g\frac{\partial D}{\partial x} \\ \frac{dv}{dt} + fu &= -g\frac{\partial D}{\partial y} \\ \frac{dD}{dt} + D\left(\frac{\partial u}{\partial x} + \frac{\partial v}{dy}\right) &= 0 \end{aligned}\right\} \qquad (5.1)$$

where u and v are the components along the x- and y-axes respectively of the horizontal wind $\mathbf{v}$, f is the Coriolis parameter, D denotes the depth of the fluid and d/dt only contains the horizontal convective accelerations. We have here made use of the hydrostatic equation. From these equations the well-known vorticity equation is derived

$$\frac{d}{dt}(\zeta + f) = -(f + \zeta)\,\mathrm{div}\,\mathbf{v} \qquad (5.2)$$

where ζ denotes the relative vorticity. In case of a fixed top on the fluid the right side is equal to zero and this simplified form of (5.2) is the basis of all barotropic forecasts published up to the present time.

Any results derived for such a model must be "translated" in terms of the real atmosphere. Thus CHARNEY (1949) has introduced the concept of the equivalent barotropic level of the atmosphere and justified the use of data at 500 mb in forecasting with (5.2). In (a) we shall give some further considerations on this problem starting from the general three-dimensional equations.

The geostrophic approximation has been of great importance for our possibilities of utilizing eq. (5.2) for forecasting purposes. In an other article in this issue of Tellus Charney discusses the more general equation relating wind and pressure in the nondivergent case. (CHARNEY 1955). Similar considerations have been made by the author and a detailed discussion of this generalization of the geostrophic relation will be given in (b).

In section (c) we shall discuss the importance of removing the restriction of zero divergence. In doing so the balance equation discussed in (*b*) becomes more complicated and some modifications of it are presented in (*d*).

It is hoped that this detailed discussion of the barotropic model shall serve the two-fold purpose of improving the present procedure in applying the barotropic model and form a starting-point for a discussion of more complicated models of the atmosphere.

a. *The equivalent barotropic atmosphere*

There exist marked differences between the real atmosphere and the barotropic nondivergent model. Thus the wind increases from relatively small values at the surface of the earth to a maximum at the tropopause level. The vertical velocity in reality seems to have a maximum somewhere in the middle of the atmosphere and is equal to zero at the surface of the earth and also in the vicinity of the tropopause, implying opposite sign of the divergence in the lower and upper part of the troposphere. A number of approximations therefore are necessary to arrive at the simplified equation used in the present computations with the barotropic model. The question arises if some of these simplifications can be removed without having to introduce several parameters. It is clear that some kind of multiple parameter model ultimately will be used in numerical forecasting but we shall here see if it is possible to make any appreciable improvements within the framework of the simple barotropic model. Let us for this purpose turn back to the complete hydrodynamic equations in three dimensions.

For this discussion we shall make use of the hydrostatic relation

$$\frac{\partial p}{\partial z} = -g\varrho \qquad (5.3)$$

with the aid of which we introduce p as our vertical coordinate. The two horizontal components of the equation of motion are combined into the vorticity equation and the divergence equation. At this stage we shall only make use of the first one. The divergence equation will be used in 5 b and 5 d to relate

the pressure and wind fields to each other. The complete vorticity equation reads

$$\frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla (\zeta + f) = -\omega \frac{\partial \zeta}{\partial p} + (f + \zeta) \frac{\partial \omega}{\partial p} + \\ + \frac{\partial \omega}{\partial y} \cdot \frac{\partial u}{\partial p} - \frac{\partial \omega}{\partial x} \cdot \frac{\partial v}{\partial p} \qquad (5.4)$$

where $\omega = dp/dt$ and where we have made use of the continuity equation. At this stage CHARNEY (1949) makes the following assumptions.

1) The variation of the wind with height is fairly well represented by

$$\mathbf{v}(x, y, p) = A(p) \bar{\mathbf{v}}(x, y) \qquad (5.5)$$

where $A(p)$ is an empirical function for which

$$\frac{1}{p_0} \int_0^{p_0} A(p) dp = 1 \qquad (5.6)$$

$\bar{\mathbf{v}}$ is the mean velocity averaged over the vertical.

2) $\zeta \ll f$ and therefore the vertical advection of vorticity, $(\omega \cdot \partial \zeta / \partial p)$, and $\zeta \cdot \partial \omega / \partial p$ are small terms in comparison with $f \cdot \partial \omega / \partial p$.

3) The turning of the vortex tubes expressed by $(\partial \omega / \partial y \cdot \partial u / \partial p - \partial \omega / \partial x \cdot \partial v / \partial p)$ can be neglected.

By integrating over the vertical Charney shows that

$$\frac{\partial \zeta^\star}{\partial t} + \mathbf{v}^\star \cdot \nabla (\zeta^\star + f) = 0 \qquad (5.7)$$

Here $\zeta^\star$ and $\mathbf{v}^\star$ represent the vorticity and velocity at a level $p^\star$ where

$$A(p^\star) = \frac{1}{p_0} \int_0^{p_0} [A(p)]^2 dp \qquad (5.8)$$

It turns out that $p^\star$ is close to 500 mb on an average. This is the justification for using the simple vorticity equation valid for an incompressible and homogeneous atmosphere for forecasting the changes at 500 mb in the real atmoshpere.

However, none of these three assumptions are valid in reality, as has been pointed out several times by other authors. The use of multiple-parameter models represent attempts to remove the restrictions introduced by these assumptions in particular the first one. We shall here only discuss the barotropic model and therefore the first assumption will be retained in spite of the fact that large errors are introduced in highly baroclinic fields.

The second assumption introduces errors, which can easily be estimated in an atmosphere with the vertical distribution of the wind given by (5.5). Consider for example a region with a positive value of ζ. Because of the general character of $A(p)$ it follows that $\partial \zeta / \partial p < 0$. Hence, in case this is a region of upward motion $\omega \cdot \partial \zeta / \partial p$ averaged through the whole troposphere is a positive quantity. In such an area of upward motion the divergence $(= -\partial \omega / \partial p)$ is positive at upper levels and negative at lower levels, but the integrated value over the vertical is close to zero. However, if ζ is positive, it is larger at higher levels than close to the surface of the earth $(\partial \zeta / \partial p < 0)$ and the average of $\zeta \cdot \partial \omega / \partial p$ is therefore also different from zero. It is easily seen that these two terms give a contribution in the same direction. We shall here give a rough estimate of the importance of this effect. As will be seen from the following derivation the result depends on the assumption of the vertical distribution of the horizontal wind and the vertical velocity. It is difficult to see what the various assumptions will mean and only numerical forecasts using the final formula will answer this question.

For simplicity we shall assume

$$A(p) = 2 \cdot \frac{p_0 - p}{p_0 - p_1} \qquad (5.9)$$

which gives $\mathbf{v}(x, y, p_0) = 0$ and $\mathbf{v}(x, y, p_1) = 2 \bar{\mathbf{v}}$. p_0 is the pressure at the surface of the earth and p_1 the pressure at the tropopause. Let us restrict the following considerations to the troposphere. (5.9) then describes average conditions quite well (BOLIN, 1953a).

It has already been mentioned that the vertical velocity seems to have an extremum somewhere in the middle of the troposphere. This is mainly a result of the existing vertical distribution of the wind. If it is accepted that the vertical variation of the wind can be described fairly well with the expression (5.9) the atmosphere must contain processes that try to maintain such a distribution. This means that $\partial \zeta / \partial t$ in (5.4) also must be a function of p approximately represented by $A(p)$. As-

suming for a moment that $\zeta \ll f$ and (5.4) can be written

$$\frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla (\zeta + f) = f \cdot \frac{\partial \omega}{\partial p} \qquad (5.10)$$

It is then easy to show that $|\omega|$ must have a maximum in the middle of the troposphere to have this equation satisfied at all levels between $p = p_0$ and $p = p_1$. For the case that $\omega(p_0) = \omega(p_1) = 0$ one obtains

$$\omega = \frac{\overline{\mathbf{v}} \cdot \nabla \bar{\zeta}}{3f} \Delta p (\pi^3 - \pi^2 - \pi + 1) \qquad (5.11)$$

where

$$\pi = \frac{2p - p_1 - p_0}{p_0 - p_1}; + 1 \leq \pi \leq -1 \qquad (5.12)$$

and $\Delta p = \frac{1}{2}(p_0 - p_1)$. In view of the fact that $f \cdot \partial\omega/\partial p$ is the most important term containing ω in (5.4) one may draw the conclusion: ω has an extremum in the vicinity of the mid-troposphere which is the result of the fact that the wind increases approximately linearly with height. This forms the basis for the assumption of a parabolic variation of the vertical velocity with height implicit in most two-parameter models (cf. EADY, 1952; ELIASSEN 1952).

It is quite obvious that any assumption of the form (5.5) means an inconsistency because the equations governing the motion of the atmosphere are non-linear in their character, which means a continuous activation of higher modes also along the vertical. Instead of requiring that the vorticity equation is satisfied at all levels we shall therefore follow the procedure used by EADY (1952) and ask for the vorticity equation to be satisfied if integrated over the vertical in various ways. Let us for simplicity assume a parabolic distribution of ω with p instead of (5.11), since the detailed structure is of no great importance in estimating the other terms depending upon ω in the vorticity equation.

$$\omega = 4\omega_m \frac{(p - p_1)(p_0 - p)}{p_0 - p_1} \qquad (5.13)$$

It is here also assumed that $\omega(p_0) = \omega(p_1) = 0$, an assumption which will be discussed later (5 c). Introducing (5.9) and (5.13) into (5.4) and integrating between p_0 and p_1, we obtain

$$\frac{\partial \bar{\zeta}}{\partial t} + \frac{4}{3} \overline{\mathbf{v}} \cdot \nabla \bar{\zeta} + \overline{\mathbf{v}} \cdot \nabla f = \frac{4}{3} \frac{\omega_m \bar{\zeta}}{\Delta p} \qquad (5.14)$$

The turning of the vortex tubes has been neglected for the time being. Another relation is obtained in the same way as by EADY (l.c.). Instead of adding the contribution from each half of the atmosphere we subtract them. Thus

$$\frac{\partial \bar{\zeta}}{\partial t} + 2\overline{\mathbf{v}} \cdot \nabla \bar{\zeta} + \overline{\mathbf{v}} \cdot \nabla f = 2 \frac{\omega_m}{\Delta p} (f + \bar{\zeta}) \qquad (5.15)$$

From (5.14) and (5.15) we derive the following expression for the vertical velocity

$$\omega_m = \frac{\Delta p}{3f + \bar{\zeta}} \overline{\mathbf{v}} \cdot \nabla \bar{\zeta} \qquad (5.16)$$

and the vorticity equation may be written

$$\frac{\partial \bar{\zeta}}{\partial t} + \frac{4f}{3f + \bar{\zeta}} \overline{\mathbf{v}} \cdot \nabla \bar{\zeta} + \overline{\mathbf{v}} \cdot \nabla f = 0 \qquad (5.17)$$

To be able to compare this with (5.7) we introduce $\zeta^\star = {}^4/_3 \cdot \bar{\zeta}$ and $\mathbf{v}^\star = {}^4/_3 \cdot \overline{\mathbf{v}}$. Thus

$$\frac{\partial \zeta^\star}{\partial t} + \frac{4f}{4f + \zeta^\star} \mathbf{v}^\star \cdot \nabla \zeta^\star + \mathbf{v}^\star \cdot \nabla f = 0 \qquad (5.18)$$

If $\zeta^\star \approx f$ an error of about 25 % is made in evaluating the vorticity change by using equation (5.7) instead of (5.18). The approximation made in disregarding the horizontal gradients of the vertical velocity is in most cases permissible. These terms are of course of principal importance in cases of shear lines or intense jet streams, but it is doubtful if such phenomena can be treated with a one-parameter model, which describes the vertical wind shear relatively crudely.

Equation (5.18) is interesting in that there is a systematic difference between areas of cyclonic and anticyclonic relative vorticity. The former move more slowly and the latter more rapidly than the pure advection indicates. Obviously the absolute vorticity is not conserved. However, the changes depend upon the gradient of the relative vorticity, which is small ($= -\nabla f$) in areas where $\eta^\star$ has a maximum or a minimum. Thus the total range within which $\eta^\star$ varies over the field remains approximately constant in time.

It is difficult to judge from the error maps of the series of computations presented above, if (5.18) would mean an improvement of the present method. This will be tested in coming

forecasts. It should be kept in mind, however, that it is based on the assumption that the vertical variation of wind as well as vertical velocity is the same everywhere. We know, for example, that there are systematic differences between cyclones and anticyclones noticeable for example in systematic variations in the height of the equivalent barotropic level (cf. BOLIN and CHARNEY, 1951). This may counteract the effects described.

b. *The relation between wind and pressure in the nondivergent case*

All three forecast-equations (5.2), (5.7) and (5.18) have the characteristic feature that a knowledge of the wind field at one instant permits an evaluation of the development in time without any knowledge of the pressure field. In reality the pressure field is better known than the wind field and therefore the wind observations usually are supplemented by the geostrophic wind computed from the pressure field. *The reason for introducing the geostrophic approximation in the barotropic model is therefore merely to be able to use both pressure and wind observations as initial data in the computations.* It is not necessary from the computational point of view. We shall here discuss the more general relation between wind and pressure in some detail.

Let us for this purpose consider an internally consistent model of the atmosphere. At this first stage we again choose a homogeneous and incompressible fluid with a fixed upper surface. Thus div $\mathbf{v} \equiv 0$. Taking the divergence of the equations of motion we arrive at the following equation relating the pressure field and the wind field to each other:

$$\nabla^2\phi = f\zeta + 2\left(\frac{\partial u}{\partial x}\cdot\frac{\partial v}{\partial y} - \frac{\partial v}{\partial x}\cdot\frac{\partial u}{\partial y}\right) - f_y u + f_x v \tag{5.19}$$

where $f_y = \partial f/\partial y$, $f_x = \partial f/\partial x$ and $\phi = p/\varrho$. Since the motion is assumed to be non-divergent we can introduce a stream function ψ defined by

$$\left.\begin{aligned} u &= -\frac{\partial\psi}{\partial y} = -\psi_y \\ v &= \frac{\partial\psi}{\partial x} = \psi_x \end{aligned}\right\} \tag{5.20}$$

Thus

$$\nabla^2\phi = f\nabla^2\psi + 2\left(\psi_{xx}\psi_{yy} - \psi_{xy}^2\right) + f_y\psi_y + f_x\psi_x \tag{5.21}$$

If neglecting the variation of the Coriolis parameter and the non-linear terms we obtain

$$\nabla^2\phi = f\zeta \tag{5.22}$$

which is the approximate form used in evaluating the vorticity with the aid of the geostrophic approximation. This relation is used in all barotropic (and barolinic) computations published so far. The neglections made in using (5.22) are of two different kinds. Integrating (5.21) over an area S and making use of Gauss theorem gives

$$\left.\begin{aligned} \int_L \frac{\partial\phi}{\partial n}\,dl = \iint_S f\cdot\zeta\,dS + \int_L (\mathbf{v}\cdot\nabla\mathbf{v})_n\,dl + \\ \iint_S (f_y\psi_y + f_x\psi_x)\,dS \end{aligned}\right\} \tag{5.23}$$

Here n means the normal direction to the boundary, l is the coordinate along the boundary L. From (5.21) and (5.23) we see that the non-linear term $(\psi_{xx}\psi_{yy} - \psi_{xy}^2)$ is of importance for a correct description of relatively small scale systems. Locally it may approach the size of the two terms retained in using the approximate relation (5.22). However, if integrating over a large area the positive and negative contributions to the integral approximately compensate each other. The term depending on the variation of the Coriolis parameter, on the other hand, is locally small, but has the same sign as the westerly component of the motion and may therefore have the same sign over large areas. It is of particular importance in describing the large scale systems in the atmosphere.

It is quite clear from (5.21) that ϕ is uniquely determined, if ψ is known over the area and ϕ is given on the boundaries. In reality, however, we observe the height of the pressure surfaces, i.e. the pressure field, and winds at these levels. Let us assume that the wind field at 500 mb is to be used in the non-divergent barotropic model. The problem is then to determine the stream function ψ initially so that the wind pattern thus defined and the corresponding pressure field defined by (5.21) agree with the observations as well as possible. The

present way of analysing a 500 mb map is essentially an analysis of the pressure field, using the observed winds as a guide in areas with few observations. We shall at this moment assume that it represents ϕ and our problem is to determine ψ from a given pressure field.

Looking upon (5.21) in this way it is a special case of the Monge-Ampère's differential equation and possesses certain characteristic features which are of interest here. It is of the elliptic type if

$$\nabla^2\phi - f_y\psi_y - f_x\psi_x > -\frac{f^2}{2} \qquad (5.24)$$

It then has two and only two solutions, if the boundary values of ψ are specified. To explore the character of the non-linear terms in the equation (5.21) and this criterion we shall assume that $f_x = f_y = 0$ for the time being. Thus (5.24) becomes

$$\frac{\nabla^2\phi}{f} > -\frac{f}{2} \qquad (5.25)$$

This means that the geostrophic vorticity must never be smaller than $-f/2$. In this case (5.21) may be written

$$(2\psi_{xx}+f)(2\psi_{yy}+f) - 4\psi_{xy}^2 > 0 \quad (5.26)$$

Thus the product $(2\psi_{xx}+f)(2\psi_{yy}+f)$ always is positive and either both factors are positive or both are negative. The two possible solutions are characterized by

$$\left.\begin{aligned}\zeta_1 = \nabla^2\psi_1 > -f\\ \zeta_2 = \nabla^2\psi_2 < -f\end{aligned}\right\} \qquad (5.27)$$

or

$$\left.\begin{aligned}\eta_1 > 0\\ \eta_2 < 0\end{aligned}\right\} \qquad (5.28)$$

η being the absolute vorticity. For continuity reasons it therefore follows that the solution of (5.21) is characterized by the fact that the absolute vorticity is positive everywhere or negative everywhere. In the northern hemisphere of course the only solution of interest is the one where $\eta > 0$. In the case of a circular vortex we here recognize a well-known fact. For a given pressure field there exist two possible solutions to the gradient wind equation.

If the inequality (5.25) is not fulfilled equation (5.21) is hyperbolic in this region. However, this is usually the case only in very limited areas. For such an area specifying ϕ in the interior of it and ψ on the boundaries usually means an over-determination of the problem and no solutions exist. The features of the equation in this sense are very similar to those of linear hyperbolic equations of second order.

It is not unusual to find areas on a 500-mb map within which (5.25) is not satisfied. If introducing the notations

$$\left.\begin{aligned}A = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} = -2\psi_{xy}\\ B = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} = \psi_{xx} - \psi_{yy}\end{aligned}\right\} \qquad (5.29)$$

we can write (5.21)

$$\nabla^2\phi = f\zeta - \frac{1}{2}(A^2 + B^2 - \zeta^2) \qquad (5.30)$$

(cf. SHERMAN, 1952; PETTERSSEN, 1953). We here still have assumed that $f_x = f_y = 0$. A and B are the two components of the deformation field. Whatever values of ζ are introduced in (5.30), $f\zeta + \frac{1}{2}\zeta^2 \geqslant -f^2/2$. Therefore if (5.25) is not satisfied the deformation field is of importance (or it is not permissible to make the assumption of non-divergence). *It is not possible to draw the conclusion that the absolute vorticity is negative by only consulting the pressure field (and its derivatives) in the point under consideration.*

Let us next consider the terms depending upon the variation of the Coriolis parameter. To clarify some facts regarding their importance we shall for a moment neglect the non-linear terms and thus write (5.21) as

$$\frac{1}{f}\nabla^2\phi = \nabla^2\psi + \frac{f_y}{f}\psi_y \qquad (5.31)$$

where we for simplicity have assumed that f is a function of y only. In middle latitudes the flow usually has a westerly component, i. e. $\psi_y < o$. The neglection of the term depending upon the variation of the Coriolis parameter then means an under-estimation of vorticity. Differentiating (5.31) with respect to time and making use of the vorticity equation

$$\frac{\partial\zeta}{\partial t} = \frac{\partial\zeta_g}{\partial t} + f_y\frac{\partial u}{\partial t} = -\mathbf{v}\cdot\nabla(\zeta+f) \quad (5.32)$$

ζ_g being the geostrophic vorticity $\zeta_g = f^{-1}\nabla^2\phi$. (5.32) shows that the geostrophic vorticity is

not conserved but decreases in an area were the westwind increases and vice versa. Accordingly errors are introduced in $\partial\phi/\partial t$ if neglecting the variation of the Coriolis parameter. Let us for example assume that u increases with 10 m/sec over an area of $5 \cdot 10^6$ km^2. Putting $f_y/f = 2 \cdot 10^{-7}$ m^{-1} we obtain an error in evaluating the circulation around this area which corresponds to a velocity of about 2 m/sec along the periphery. This velocity field then induces systematic errors in the vorticity advection.

The velocity field given by (5.20) is nondivergent, the geostrophic wind is not. It is of course questionable if the assumption of nondivergence is permissible, but if we want to remove this assumption it certainly must involve more than merely estimating the divergence geostrophically (cf. 5 d). It is therefore advantageous to apply this assumption strictly. The velocity field to be added to the geostrophic one to make it nondivergent is easily found by introducing the notations

$$\left.\begin{aligned} u &= -\frac{1}{f}\frac{\partial\phi}{\partial y} + \frac{1}{f}\frac{\partial\alpha}{\partial x} \\ v &= \frac{1}{f}\frac{\partial\phi}{\partial x} + \frac{1}{f}\frac{\partial\alpha}{\partial y} \end{aligned}\right\} \qquad (5.33)$$

where $$\nabla^2\alpha = f_y v$$

The intensity of this additive velocity component is of the same order of magnitude as the one mentioned in the previous paragraph.

We shall not prolong this discussion, as the importance of the different terms in (5.21) is best illustrated by making a forecast using the correctly computed stream-function.

If the equation is elliptic in its type one can solve for ψ from a knowledge of ϕ by relaxation methods. The criterion for ellipticity depends upon the solution itself (5.24), but if neglecting the variation of the Coriolis parameter it is only a function of the given values of ϕ and f, (5.25). To avoid having to investigate if this criterion is satisfied at every step in the iterative process and to investigate the importance of the various terms separately the terms depending on f_x and f were neglected in the following considerations. Let us assume we have an approximate solution ψ'. Denoting the residual by ε we obtain

$$f\nabla^2\psi' + 2(\psi'_{xx}\psi'_{yy} - \psi'^2_{xy}) - \nabla^2\phi = \varepsilon \qquad (5.34)$$

Now vary the function ψ' in order to decrease the residual ε successively. We get

$$\left.\begin{aligned} (f + 2\psi'_{yy})\,\delta\psi'_{xx} - 4\psi'_{xy}\delta\psi'_{xy} + \\ + (f + 2\psi'_{xx})\,\delta\psi'_{yy} = \delta\varepsilon \end{aligned}\right\} \qquad (5.35)$$

This is a linear elliptic differential equation in $\delta\psi'$ if (5.25) is satisfied. By using the approximate solution ψ' we can estimate this correction $\delta\psi'$ by solving this linear equation by relaxation methods. Introducing finite differences for the evaluation of the derivatives of $\delta\psi'$ (5.35) becomes

$$\left[f_{ij} + 2\left(\frac{\partial^2\psi'}{\partial y^2}\right)_{i,j}\right](\delta\psi'_{i+1,j} + \delta\psi'_{i-1,j} - 2\delta\psi'_{i,j}) +$$

$$+\left[f_{ij} + 2\left(\frac{\partial^2\psi'}{\partial x^2}\right)_{i,j}\right](\delta\psi'_{i,j+1} + \delta\psi'_{i,j-1} - 2\delta\psi'_{i,j})$$

$$-\left(\frac{\partial^2\psi'}{\partial x\partial y}\right)_{i,j}(\delta\psi'_{i+1,j-1} + \delta\psi'_{i-1,j-1} -$$

$$-\delta\psi'_{i+1,j-1} - \delta\psi'_{i-1,j+1}) = \frac{\delta\varepsilon_{i,j}}{m^2_{i,j}} \qquad (5.36)$$

where

$$m^2_{ij} = \frac{g(\varphi_{ij})}{(\Delta s)^2} \qquad (5.37)$$

Δs being the gridsize at a particular latitude. $g(\varphi_{ij})$ is a function of the latitude and depends upon the map projection used. At the point (i,j) (5.36) may be satisfied by putting

$$\delta\psi'_{ij} = -\frac{\delta\varepsilon_{ij}}{4m^2_{ij}(f_{ij} + \nabla^2_{ij}\psi')} \qquad (5.38)$$

and

$$\left.\begin{aligned} \delta\psi'_{i+1,j+1} = \delta\psi'_{i,j+1} = \delta\psi'_{i-1,j+1} = \delta\psi'_{i+1,j} = 0 \\ \delta\psi'_{i-1,j} = \delta\psi'_{i+1,j-1} = \delta\psi'_{i,j-1} = \delta\psi'_{i-1,j-1} = 0 \end{aligned}\right\} \qquad (5.39)$$

If now assuming that $\delta\varepsilon_{i,j} = -\varepsilon_{i,j}$ and that $\varepsilon_{i,j}$ is given by the finite difference form of (5.34) we have improved the solution in the point (i,j) and the new value $\psi''_{i,j}$ is

$$\psi''_{i,j} = \psi'_{i,j} + \delta\psi'_{i,j} \qquad (5.40)$$

The correction thus obtained does not give a complete agreement for the new values at the point (i,j) because of the non-linear character

of (5.34) and the linearization that is involved in the derivation. It is, however, no point in trying to correct for this now before the other points have been relaxed in the same way, because it is still only an approximation. Thus we proceed to the next point and in correcting the ψ'-value we make now use of the new and better value ψ'' in the points already treated. The method then becomes similar to the so called Liebmann-method. The convergence of an iterative process of this kind has been discussed by NIRENBERG (1953).

It was mentioned before that the criterion for ellipticity usually is not fulfilled everywhere on an ordinary 500-mb map. For example on October 2, 1954, 0300 GMT the inequality (5.25) was satisfied over about 94 % of the total area. The procedure outlined above breaks down in the cases when the equation becomes hyperbolic at some point. To be able to get an idea of the importance of the term it was therefore decided to change the value of $\nabla^2\phi$ in those points so that

$$\frac{\nabla^2\phi}{f} = -\alpha\frac{f}{2} \qquad (5.41)$$

where $\alpha = 0.97 < 1$. This is of course not going to be the ultimate method to be used, but will still be of some interest here. The most serious consequence of this modification of the data is that the circulation around the total area is changed somewhat.

The results of some forecasts using the stream function thus obtained will be reported in a following issue of Tellus. Some questions regarding the transformation to a stereographic projection will also be discussed.

c. *The effect of a non-vanishing divergence*

In the derivation of (5.18) as well as the ordinary forecast equation (5.7) it was assumed that $\omega(p_0) = \omega(p_1) = 0$, i.e. the horizontal divergence integrated over the vertical is equal to zero. In other words, we have used the non-divergent barotropic model. Also the considerations in 5 b apply to this case. The fact that the pressure changes at the surface of the earth only amount to a few per cent of the total pressure indicates that this assumption is permissible as a first approximation as shown by CHARNEY (1949). Extending the integration to the top of the atmosphere ($p_1 = 0$) he derived the expression

$$\frac{\partial\bar{\zeta}}{\partial t} + \mathbf{v}\cdot\nabla(K\bar{\zeta}+f) = \frac{f}{p_0}\frac{\partial p_0}{\partial t} \qquad (5.42)$$

where p_0 is the surface pressure. (Here again it has been assumed $\zeta \ll f$).

Even if this term is small it is of principal importance as has been shown by ROSSBY (1945) and YEH (1949). For such a demonstration it is again feasible to study an idealized model of the atmosphere, i.e. a homogeneous and incompressible fluid, however, now with a free surface to permit the divergence to be different from zero. They found that the divergence term determines the character of dispersion in such fluid and Yeh also attempted to account for the existence and the permanence of blocking waves. The importance of this term is mainly noticeable for very long waves.

The question is, however, if this model takes the divergence into consideration in the way it appears in the real atmosphere. We are interested in studying the changes of the flow-pattern within the troposphere and the vertical integration should therefore be extended over the troposphere only. We now know that the vertical velocities at the tropopause are quite systematic in that cyclones have a low tropopause and anticyclones a high one. These variations easily amount to 200 mb indicating that *the horizontal divergence or convergence within the troposphere as a whole may change the depth of the troposphere with values up to about 20 %*. This effect is approximately depicted by modifying the barotropic model in the following way: On top of the homogeneous and incompressible fluid layer, which is supposed to represent the troposphere, and is in motion accordingly, we place another fluid layer at rest with a density chosen in such a way that the vertical motion of the interface is of the same relative magnitude as the motion of the tropopause in reality. The motion induced in the upper fluid can be assumed to be small by letting the depth of it approach infinity. The motion in the lower fluid then is governed by the equations

$$\left.\begin{aligned}\frac{du}{dt} - fv &= -\varkappa g\frac{\partial D}{\partial x}\\ \frac{dv}{dt} + fu &= -\varkappa g\frac{\partial D}{\partial y}\end{aligned}\right\} \qquad (5.43)$$

$\varkappa = (\varrho - \varrho')/\varrho$, where ϱ and ϱ' are the densities of the lower and upper fluid respectively. The vorticity equation (5.2) is not changed but we are not permitted to put the divergence equal to zero. We have instead to relate it to the changes of the depth of the fluid as given by the continuity equation in (5.1). To get a crude idea of the importance of the divergence term we shall first make the same approximations as used in deriving the forecast equation (5.7) except for putting div $\mathbf{v} = 0$. This is of course inconsistent in certain respects. We shall return to a more exact formulation of the problem in section (d). For the time being we thus introduce the geostrophic approximation

$$\zeta = \frac{\varkappa g}{f} \nabla^2 D \qquad (5.44)$$

The variation of f with latitude is neglected. For a given distribution of ζ we see that the total range for D is inversely proportional to $\varkappa$. For a typical wind pattern at 500 mb the total range of the height of the 500 mb surface is 600—800 m. The variations in the height of the tropopause amount to five or ten times this value. Changes of this order of magnitude at the interface between the two fluids in our model are obtained by putting $\varkappa \approx 8$. From (5.43) we now can derive the forecast equation

$$(\nabla^2 - \lambda^2) \frac{\partial D}{\partial t} = J(\eta, D) \qquad (5.45)$$

where

$$\lambda^2 = \frac{f^2}{\varkappa g D_0} \qquad (5.46)$$

It has here been assumed that the convective terms in the continuity equation can be neglected, which is true in case of exact geostrophic flow. The two terms on the left side of the equation (5.45) become equally important when the wave length L is given by $L = 2\pi\sqrt{\varkappa g D}/f^2$. For the values $\varkappa = 1/8$, $g = 10$ m sec^{-2}, $D_0 = 10^4$ m and $f = 10^{-4}$ sec^{-1} we obtain $L = 7{,}000$ km.

Yeh (1949) has studied the properties of waves in a barotropic model with a *free* surface and his results are directly transferable to this model provided (5.46) is used for evaluating λ. This modification is, however, very important. For example in applying the results to a study of blocking Yeh is forced to assume the solitary wave to be of very large scale to obtain significant results. The considerations above reduce the scale by a factor of about $\sqrt{\varkappa} \approx 3$ which brings the systems in very good agreement with the size of actually observed blocking ridges. This supports the idea that already the simple equation (5.45) would mean an improvement compared with the present forecast formula (5.7). Only the largest components of the flow are influenced by this change but in view of the fact that the results of the computations presented in section 3 seem to indicate that (5.7) describes the development of the very largest systems less accurately than middle sized disturbances it would be of interest to use eq. (5.45) in some actual forecasts. This is being planned.

d. *The interplay between the wind and pressure fields in the divergent case*

The treatment in the previous section was very approximate. The purpose was only to point out one principle effect of a non-vanishing divergence. It is now of some interest to extend the reasoning and to investigate in which way the more general divergence equation also here might replace the geostrophic approximation. In this way we shall arrive at a more general set of forecast equations for the barotropic model. Some comments on this problem have also been given by Charney (1955).

We shall make use of the following very important fact concerning the average motion of the atmosphere: *The horizontal divergence is one order of magnitude smaller than the relative vorticity*. This is obviously true if considering the average through the whole atmosphere, since the percentage variation of the pressure at the surface is small (10^{-6}sec^{-1}). Even if we only consider the troposphere and assume the tropopause to be a material surface, we find that the height variations of the tropopause in connection with disturbances in the troposphere give rise to an average horizontal divergence in the troposphere which is less than 10^{-5}sec^{-1}. The relative vorticity, on the other hand, is abont one order of magnitude larger.

The discussion here will be limited to the barotropic case and we shall use the model of a homogeneous and incompressible atmosphere with an other infinitely deep upper layer on

top. The motion of such a fluid is governed by the equations (5.1) if replacing g by $\varkappa g$ where as before $\varkappa = (\varrho - \varrho')/\varrho$. ϱ and ϱ' are the densities of the lower and upper fluids respectively.

From the equations (5.1) we derive the vorticity equation and the divergence equation and obtain the following set of equations.

$$\frac{\partial \zeta}{\partial t} + \mathbf{v} \cdot \nabla (f+\zeta) + (f+\zeta) \operatorname{div} \mathbf{v} = 0 \quad (5.47)$$

$$\frac{\partial}{\partial t}(\operatorname{div} \mathbf{v}) + \mathbf{v} \cdot \nabla (\operatorname{div} \mathbf{v}) + (\operatorname{div} \mathbf{v})^2 - \zeta f - 2\left(\frac{\partial u}{\partial x} \cdot \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} \cdot \frac{\partial v}{\partial x}\right) + f_y u - f_x v + g \nabla^2 D = 0 \quad (5.48)$$

$$\frac{\partial D}{\partial t} + \mathbf{v} \cdot \nabla D + D \operatorname{div} \mathbf{v} = 0 \quad (5.49)$$

Putting div $\mathbf{v} = 0$ the vorticity equation shows that the wind field completely determines its own development. The divergence equation is reduced to a relation between the pressure and wind fields not containing any time derivatives and the continuity equation becomes an identity.

A non-vanishing divergence influences the vorticity equation in two ways. The wind can no longer be represented by a stream function only, but the divergence field must be associated with a velocity field, which can be described by a velocity potential. Let us therefore put

$$\left.\begin{aligned} u &= -\frac{\partial \psi}{\partial y} + \frac{\partial \chi}{\partial x} \\ v &= \frac{\partial \psi}{\partial x} + \frac{\partial \chi}{\partial y} \end{aligned}\right\} \quad (5.50)$$

where

$$\left.\begin{aligned} \nabla^2 \psi &= \zeta \\ \nabla^2 \chi &= \operatorname{div} \mathbf{v} \end{aligned}\right\} \quad (5.51)$$

Secondly, the absolute vorticity is not conserved because of the divergence. However, since the vorticity is one order of magnitude larger than the divergence and the time-scale of atmospheric disturbances is one day or more both these effects are comparatively small. The major changes of the flow are already obtained by the non-divergent model as indicated by the comparatively successful results with this model. *It is therefore not necessary to know the divergence field with the same percentage accuracy as the vorticity field.* We therefore are permitted to neglect some of the terms in the equations (5.48) and (5.49), since these equations are used for obtaining the proper relation between the divergence field and the vorticity field. We shall later return to the question of the conditions under which these simplifications are permissable, as well as their physical meaning. Neglecting small terms in (5.48) we obtain the same balance equation between the pressure field and the wind field as in the non-divergent case.

$$g \nabla^2 D = \zeta f + 2\left(\frac{\partial u}{\partial x} \cdot \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} \cdot \frac{\partial v}{\partial x}\right) - f_y u + f_x v \quad (5.52)$$

while (5.49) will be used without any simplifications. The following procedure for numerical integration then is suggested: At a certain time $\tau \cdot \Delta t$ we know $\psi = \psi^\tau$, $\chi = \chi^\tau$ and $D = D^\tau$ and similarly at previous times. (Δt is the time step used in the integration.) Thus $(\partial \zeta / \partial t)^\tau$ can be determined from (5.47) and $\zeta^{\tau+1}$ is obtained by linear extrapolation

$$\zeta^{\tau+1} = \zeta^{\tau-1} + 2\left(\frac{\partial \zeta}{\partial t}\right)^\tau \Delta t \quad (5.53)$$

We then obtain $\psi^{\tau+1}$ by solving the Poisson equation (5.51) relating ψ and ζ. This is done by relaxation method and the linear extrapolation from $\psi^{\tau-1}$ and ψ^τ gives a first guess. $\psi^{\tau+1}$ determines the non-divergent component of the wind field. We next want to determine the corresponding pressure field $D^{\tau+1}$ with the aid of (5.52). The major part of the wind field $\mathbf{v}^{\tau+1}$ is given by $\psi^{\tau+1}$ and the small component given by $\chi^{\tau+1}$ can be neglected, since we only need approximate values of D and still are able to forecast ψ with good accuracy. Thus the right side of (5.52) can be evaluated and $D^{\tau+1}$ is obtained by solving a Poisson equation for example by relaxation. Finally we want to get a better estimate of $\chi^{\tau+1}$ to be used in the next time-step. The wind and pressure fields and their changes imply a certain divergence field according to (5.49). Putting

$$\left(\frac{\partial D}{\partial t}\right)^{\tau+1/2} = \frac{1}{\Delta t}(D^{\tau+1} - D^\tau) \quad (5.54)$$

and evaluating $\mathbf{v}^{\tau+1/2} \cdot \nabla D^{\tau+1/2}$ as the mean value over the time interval between $\tau \cdot \Delta t$ and $(\tau+1)\Delta t$ we can evaluate $(\operatorname{div} \mathbf{v})^{\tau+1/2}$. With the aid of (5.51) we can determine $\chi^{\tau+1/2}$ and obtain finally $\chi^{\tau+1}$ by extrapolation

$$\chi^{\tau+1}=\chi^{\tau}+2\,(\chi^{\tau+1/2}-\chi^{\tau}) \qquad (5.55)$$

Thus we are ready for the next time step. Since we only need an approximate value of χ it may even be permissable to neglect the convective term i (5,49) as it is small in quasi-geostrohic flow.

At the initial time $\tau = 0$ only D° and ψ° are known, the latter from the solution of the simplified divergence equation as outlined in the previous section. We therefore first have to determine χ°, which can be done approximately by neglecting the influence of the divergence in (5,47). Thus we can get an approximate value of $(\partial\zeta/\partial t)^\circ$ and thus also of $(\partial\psi/\partial t)^\circ$. With the aid of (5.52) we can evaluate $(\partial D/\partial t)^0$ and then solve for χ^0 from (5.49). The computations can then start as described previously except for the fact that uncentered differences have to be used.

The scheme outlined here is possible only if ψ represents the major part of the wind field and the wind field corresponding to χ merely is a comparatively small correction.

Forecasts with this method are being prepared and the results will be reported in a following issue of Tellus. The details in the computational procedure will then also be given.

We shall give a few additional comments on the approximations made. The most serious one is the neglection of the time-dependant term $d(\text{div}\,\mathbf{v})/dt$ in (5.48). Hereby all gravity-inertia oscillations are eliminated and a balanced state between pressure and wind is assumed. We can get an approximate idea of the importance of these processes in the following way:

Since (div $\mathbf{v}$) is small, (div $\mathbf{v}$)^2 certainly can be neglected in (5.48). Furthermore we replace the total derivative $d(\text{div}\,\mathbf{v})/dt$ by the local one neglecting the convective term $\mathbf{v}\cdot\nabla\,(\text{div}\,\mathbf{v})$. We shall see that maintaining the local time derivatives in the equation permits the existance of gravity-inertia waves the speed of which (c) is about 100 m/s. As long as c is large compared with $\mathbf{v}$ these waves are fairly well described even if neglecting the convective term. Under all circumstances it means less simplifications than merely using the balance equation (5.52). We shall furthermore disregard the convective term in the continuity equation. This term is zero if the wind is parallell with the isobars and therefore as long as the departures from geostrophic or gradient flow are small this simplification is permissable. Finally we assume that the percentage variations of D are small and replace D by D_0. The continuity equation becomes

$$\frac{1}{D_0}\cdot\frac{\partial D}{\partial t}+\text{div}\,\mathbf{v}=0 \qquad (5.56)$$

With these simplifications combination of (5.48) and (5.56) for elimination of $\partial D/\partial t$ gives

$$\frac{\partial^2}{\partial t^2}(\text{div}\,\mathbf{v})-\varkappa g D_0\nabla^2(\text{div}\,\mathbf{v})+f\,(f+\zeta)\,\text{div}\,\mathbf{v}=$$
$$=\mathbf{v}\cdot\nabla\,(\zeta+f)+\frac{\partial}{\partial t}\left[2\left(\frac{\partial u}{\partial x}\cdot\frac{\partial v}{\partial y}-\frac{\partial u}{\partial y}\cdot\frac{\partial v}{\partial x}\right)-\right.$$
$$\left.-f_y u+f_x v\right] \qquad (5.57)$$

Let us assume that the changes of the wind field are known over a certain time interval. We can then consider (5.57) as a wave-equation with a forcing term which is a function of these variations of the wind field and div $\mathbf{v}$ is determined if proper initial conditions are prescribed. It is obvious that this viewpoint is possible only if the divergence is small so that the right side of equation (5.57) is practically independant of it. In other words the "feed back" on the forcing term from the divergence can be disregarded. The homogeneous part of this equation is the same equation (generalized to two dimensions in space) as CAHN (1945) obtained for the adjustment of the pressure and wind fields to each other. The solution of (5.57) is composed of two parts, one being the solution of the homogeneous equation with certain initial values of div $\mathbf{v}$ and $\partial\,(\text{div}\,\mathbf{v})/\partial t$ and another being the solution of the inhomogeneous equation with the initial conditions div $\mathbf{v}=\partial\,(\text{div}\,\mathbf{v})/\partial t=0$. The solution of the homogeneous equation is in principle the same as given by Cahn and takes care of the adjustment of an initial out of balance between the pressure and wind fields. The other part of the solution gives the steady adjustment of the pressure field because of the changing wind field.

CHARNEY (1955) points out that the use of the geostrophic approximation as initial conditions in the primitive equations of motion gives rise to large oscillations. It is clear from (5.57) that this must be the case. The geostrophic wind

has a small divergence $\operatorname{div} \mathbf{v}_g = v \cdot a^{-1} \cot \varphi$, where v is the wind component south—north and a is the radius of the earth. Neglecting $(\operatorname{div} \mathbf{v})^2$ and $\mathbf{v} \cdot \nabla \operatorname{div} \mathbf{v}$ in (5.48) as before we furthermore obtain at the initial instant

$$\frac{\partial}{\partial t}(\operatorname{div} \mathbf{v}) = 2\left(\frac{\partial u_g}{\partial x} \cdot \frac{\partial v_g}{\partial y} - \frac{\partial u_g}{\partial y} \cdot \frac{\partial v_g}{\partial x}\right) = \\ = \nabla \cdot (\mathbf{v}_g \cdot \nabla \mathbf{v}_g) \qquad (5.58)$$

This is also different from zero and has large values in areas of strong curvature of the flow, or if the deformation field is strong. The solution of the homogeneous equation then is a large amplitude oscillation and is a reflection of the adjustment of the wind and pressure fields towards an equilibrium characterized by (5.52). Relating the wind and pressure fields to each other by (5.52) initially means that we put $\operatorname{div} \mathbf{v} = \partial (\operatorname{div} \mathbf{v})/\partial t = 0$ and no oscillations of this kind appear.

The equation (5. 52) has a maximum speed of progagation of influences which is equal to $c_0 = \sqrt{\varkappa \cdot g D_0}$. With the values previously assigned to $\varkappa$ and D_0 we get $c_0 = 400$ km h^{-1}. A change of the wind field at one point does not influence the pressure field beyond a distance of about 600 km in $1^1/_2$ hour which is the time step used in most forecasts presented here. The use of the balance equation (5.52) on the other hand means that the pressure immediately is influenced over the whole area considered.

In the derivation of (5.57) we made some approximations in the continuity equation which were not necessary if using the balance equation (5.52). It is therefore questionable if any improvements would be obtained by using this equation. However, it would be desirable to investigate this in some more detail.

The discussion has here been restricted to a barotropic model of the atmosphere. Since c_0 is so large and the divergence small in this model, the refinements in this last section will probably not improve the forecasts essentially. It is, however, of some interest to have these processes clarified in this comparatively simple case, as the importance of gravity-inertia waves probably is greater in the baroclinic case. Here we the encounter other difficulties. The speed of the internal waves, that may exist, is smaller than the influence speed given above (cf. BOLIN 1953 b). Furthermore ROSSBY (1938) already pointed out that in the case of a one layer fluid the wind field is very little changed during the adjustment, but in a two layer fluid a considerable loss of kinetic energy takes place. This is an indication of the fact that in a stratified fluid the adjustment of a velocity field varying along the vertical is associated by a considerable redistribution of the mass field. The horizontal divergence is then not so small compared with the vorticity any longer and the procedure outlined above becomes less accurate. It is then questionable if any balance equation between the wind and the pressure fields will describe the actual processes sufficiently well.

REFERENCES

BOLIN, B., 1953 a: Multiple-parameter models of the atmosphere for numerical forecasting purposes. *Tellus* **5**, pp. 207—218.

— 1953 b: The adjustment of a non-balanced velocity field towards geostrophic equilibrium in a stratified fluid. *Tellus* **5**, pp. 373—385.

BOLIN, B., CHARNEY, J., 1951: Numerical tendency computations from the barotropic vorticity equation. *Tellus*, **3**, pp. 248—257.

CAHN, A., 1945: An investigation of the free oscillations of a simple current system. *J. of Meteorology*, **2**, pp. 113—119.

CHARNEY, J., 1949: On the physical basis for numerical prediction of large-scale motions in the atmosphere. *J. of Meteorology*, **6**, pp. 371—385.

CHARNEY, J., 1955: The use of the primitive equations of motion in numerical forecasting. *Tellus* **7**, pp. 22—26.

CHARNEY, J., PHILLIPS, N., 1953: Numerical integration of the quasi-geostrophic equations of motion for barotropic and simple baroclinic flows. *J. of Meteorology*, **10**, pp. 71—99.

CLAPP, PH. F., 1953: Application of barotropic tendency equation to medium range forecasting. *Tellus* **5**, pp. 80—94.

COURANT, R., und HILBERT, D., 1931: *Methoden der mathematischen Physik*, Berlin.

EADY, E., 1952: Note on weather computing and the so-called $2^1/_2$ dimensional model. *Tellus* **4**, pp. 157—167.

ELIASSEN, A., 1952: Simplified models of the atmosphere, designed for the purpose of numerical weather prediction, *Tellus* **4**, pp. 145—156.

FJØRTOFT, R., 1952: On a numerical method of integrating the barotropic vorticity equation. *Tellus* **4**, pp. 179—194.

NIRENBERG, L., 1953: On non-linear elliptic partial differential equations and hölder continuity. *Communications on pure and applied mathematics*, **6**, pp. 103—156.

PETTERSSEN, S., 1953: On the relation between vorticity deformation and divergence and the configuration of the pressure field. *Tellus* **5**, pp. 231—238.

ROSSBY, C.-G., 1938: On the mutual adjustment of pressure and velocity distribution in aertain simple current systems. II. *J. of Marin. Res.*, **1**, 239—263.

— 1945: On the propagation of frequencies and energy in certain types of oceanic and atmospheric waves. *J. of Meteorology*, **2**, pp. 187—204.

— 1949: On a mechanism for the release of potential energy in the atmosphere. *J. of Meteorology*, **6**, pp. 163—180.

ROSSBY, C.-G., and collaborators, 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action, *J. of Marine Research*, **2**, pp. 38—55.

SCHERHAG, R., 1948: *Wetteranalyse und Wetterprognose*, Berlin, Springer-Verlag. 424 pp.

SHERMAN, L., 1952: On the scalar-vorticity and horizontal divergence equations. *J. of Meteorology*, **9**, pp. 359—366.

STAFF MEMBERS, INSTITUTE OF METEOROLOGY UNIVERSITY OF STOCKHOLM, 1954: Results of forecasting with the barotropic model on an electronic computor (BESK), *Tellus* **6**, pp. 139—149.

YEH, T. C., 1949: On energy dispersion in the atmosphere. *J. of Meteorology*, **6**, pp. 1—16.

On the Electric Field Theory of Magnetic Storms and Aurorae

By H. ALFVÉN, The Royal Institute of Technology, Stockholm

Manuscript received November 4, 1954

Abstract

The electric field theory is discussed and further developed. The basic assumption of the theory is that the sun emits beams of rarified ionized gas which are electrically polarized. The electric field of the beam produces a discharge of Malmfors-Block's type around the earth. At the same time the electric field produces a variation in the Cosmic Ray intensity.

The theoretical treatment starts by an analysis of the motion of charged particles in the equatorial plane of the earth under the action of the electric field. This phenomenon is supposed to be of basic importance. The motion produces space charge in certain regions. This charge is supposed to leave the equatorial plane along the magnetic lines of force. These two basic assumptions, which are supported by the model experiments, lead to a theory which gives a fairly good picture of magnetic storms and aurorae.

In the earlier presentation of the theory the inertia of the beam was neglected. The theory gave a description of the main phase of a storm. The motion in the equatorial plane produces a clockwise ring current at about 7 earth-radii from which a discharge to the auroral zones takes place.

In the present paper also the inertia of the beam is taken into account. The result is that a second ring current is produced, which flows anti-clockwise at about 30 earth-radii. This current which seems to account for the *initial* phase of a storm, should be associated with an "*inner auroral zone*" with a polar distance of 5°—10°. The importance of observing aurorae and magnetic disturbances in this region is stressed.

Introduction

Like Chapman-Ferraro's theory the electric field theory of magnetic storms and aurorae is founded on Schuster's idea that the cause of these phenomena is a beam of ionized gas emitted from the sun, but this is almost the only point of agreement. The assumptions about the most important properties of the beam are different, and so are the descriptions of the phenomena which occur when the beam reaches the earth.

Concerning the beam Chapman-Ferraro's theory assumes that the *magnetic field inside the beam* is zero. This means that before the beam has reached the neighbourhood of the earth, the electric field in the beam is zero (apart from fields produced by diffusion and similar effects). Contrary to this the electric field theory assumes that the beam possesses a magnetic field H which is "frozen in" into the highly conducting matter, and that due to the motion v of the beam this magnetic field produces an electric field $E = -\frac{1}{c} v \times H$. Although it is possible that H and E have a complicated structure, only the simple case when E is homogeneous and tangential to the earth's orbit has been discussed.

The assumption that the magnetic field inside the beam is zero was natural at the time when Chapman and Ferraro made it. Since then, however, it has become evident that magneto-hydrodynamic phenomena are much more important in cosmical physics than assumed earlier, and today it seems very artificial to ignore the magnetic field. Indeed, it implies that the beam is produced in a region of the solar atmosphere where the field is zero.

The idea of an electric field associated with the beam has recently got much support from cosmic ray results (BRUNBERG and DATTNER 1954). In fact, it seems possible to explain the changes in Cosmic Ray intensity associated

with magnetic storms as produced by the electric field of the beam, whereas this effect seems to be difficult to explain according to Chapman-Ferraro's theory.

When the beam approaches the earth it is braked by the earth's magnetic field and a system of currents is produced. In Chapman-Ferraro's theory the currents are produced by the inertia of the beam, and the current system which results is identified with the currents during the initial phase. No clear picture of the main phase is given. In some papers it is claimed that the main phase is due to a ring current of Birkeland's type. In other papers the whole current system is supposed to be localized in the upper atmosphere.

In the electric field theory the electric field of the beam is supposed to be its most important property and a magnetic storm is considered as the result of the application of a field to the environment of the earth. In the earlier development of the electric field theory the main phase of a storm was described but no explanation of the initial phase was given. According to the theory the current system during the main phase consists of an *eccentric* ring current in the equatorial plane, produced by particles moving in trochoidal orbits. This current is closed by currents along the lines of force and in the auroral zones.

The purpose of the present paper is to reconsider the electric field theory. The inertia of the beam which was neglected in the earlier presentations of the theory is included in the present paper. The result of this comes out to be that the theory describes not only the main phase but also the *initial phase*. Another result of the theory is the prediction of an "*inner auroral zone*" inside the normal auroral zone.

§ 2. *Basic assumptions*

The simplest assumption we could make is that *a magnetic storm is produced by a homogeneous electric field* in the environment of the earth.

We represent the earth's field by the field from a dipole a pointing in the $-z$ direction. The beam carries a magnetic field H_0 with it, and we must make an assumption about the direction of this field. We assume that it is parallel to the z-axis, and that it is homogeneous in the relatively small region around the earth, which we are considering. Hence the total field in the equatorial plane is

$$H = H_0 + aR^{-3} \tag{1}$$

Suppose that the velocity vector v_0 of the beam points in the $-y$ direction. This means that the sun is far away in the $+y$ direction. The electric field E of the beam is

$$\vec{E} = -\frac{1}{c}\vec{v_0} \times \vec{H_0} \tag{2}$$

This means that the electric field of the beam points in the x-direction. We assume that E is constant over the whole region around the earth.

The assumptions about H_0 and E which we have made are the same as made in the theory of storm variations of the Cosmic Radiation (Alfvén, 1949, 1950, 1954).

We shall later (in § 6) make a detailed discussion of the motion of charged particles of the beam in the magnetic field from (1) under the influence of the electric field (2), but first we shall discuss the problem qualitatively.

As in most discharges electrons are the most important carriers of electric current it is reasonable to start by discussing the motion of the electrons in the beam. As the energy is much below Cosmic Ray energies, the radius of curvature of their paths is small compared to the distance to the earth. This means that they will move in trochoidal orbits and their motion can be treated by the methods outlined in Cosmical Electrodynamics Chapter II (Alfvén, 1950). Due to the electric field E their motion in the equatorial plane will be a drift in the $-y$ direction. Near the dipole this drift is superimposed on an anti-clockwise drift due to the inhomogeneity of the magnetic field. The result is that the electrons in the beam will drift in orbits as shown in Fig. 1. Superimposed on this motion there may be oscillations through the equatorial plane.

§ 3. *The ring current*

It is immediately seen that this motion is equivalent to a ring current in the equatorial plane. As the electrons move anti-clockwise, the current goes clockwise. This current gives a decrease in the magnetic field in the equatorial plane near the dipole. The electrons go closer to the dipole in the $-x$ direction (which corresponds to 18^h in the afternoon if the y-axis points towards the sun). Hence the decrease in magnetic field should be a maximum

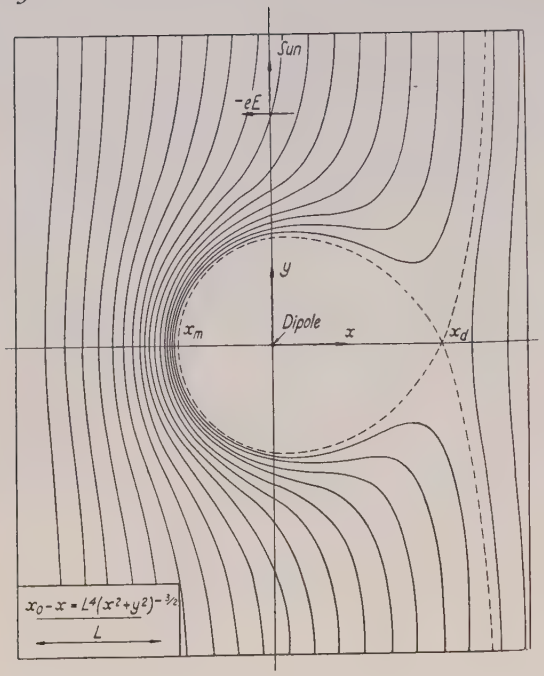


Fig. 1. Motion of the electrons in the equatorial plane of a magnetic dipole field on which is superimposed a homogeneous magnetic field, under the action of a homogeneous electric field.

at that time. This agrees very well with the equatorial disturbance during the main phase of a magnetic storm. (This can be seen by adding D_s and S_D. The general decrease in D_s during the main phase is superimposed by a negative value of S_D at 18^h and by a positive value at 6^h.)

It is important to note that in this case the ring current consists of electrons moving in trochoidal orbits. It is well known that these orbits are perfectly stable. It is of interest to compare this current with the ring current in Chapman-Ferraro's theory. In order to explain the decrease of the equatorial field during the main phase, Birkeland and Störmer supposed that there is a ring current in the equatorial plane at a few times the earth's radius. This conception has been included in Chapman-Ferraro's theory. Without any detailed arguments for it they assume that when the ionized cloud is stopped at some distance from the earth, charged particles are emitted in such a way that they move in circles around the earth, thus producing a ring current.

It has never been proved that a ring current of this type is stable. From the betatron theory we know under what conditions a particle moving in a circle with radius R in a magnetic field H has a stable orbit: If

$$H = cR^n$$

where c is a constant, the condition for stability is

$$0 > n > -1$$

As in the dipole field of the earth we have

$$n = -3$$

a circular orbit is certainly not stable. This holds for a single particle. It is very likely that it also holds for a multitude of particles. Until the contrary is proved it is reasonable to assume that a ring current of this type cannot exist.

§ 4. *The auroral zone*

Returning to the electric field theory, it is of interest that there is a certain "forbidden region" around the dipole which the electrons cannot reach. The electron density is a maximum near the border of the forbidden region. At the same time as the electrons drift in orbits of Fig. 1 they may oscillate perpendicular to the equatorial plane. When doing so they move along the magnetic lines of force (see Cosmical Electrodynamics Chapter II). If the amplitude of these oscillations is large enough they may hit the surface of the earth. As the electron density is a maximum near the border of the forbidden region it is important to see where electrons starting from this line hit the earth. This is found by projecting the borderline along the lines of force upon the earth's surface. The line which is obtained in this way has the shape shown in Fig. 2. This line is identified with the *auroral zone*. It goes closest to the pole at 6^h and its polar distance is a maximum at 18^h. The equation of the curve is given in Cosmical Electrodynamics, p. 182.

The polar distance of this curve varies with the time angle and it is important to note that the variation is of the same type as the observed time variation of the polar distance of the magnetic polar disturbance and of the aurora.

§ 5. *Terrella experiments*

We have seen that already the simple case we have discussed gives certain features, which

are characteristic to a magnetic storm: viz. an equatorial ring current and a region of disturbances which may be identified with the auroral zone. The question is whether these

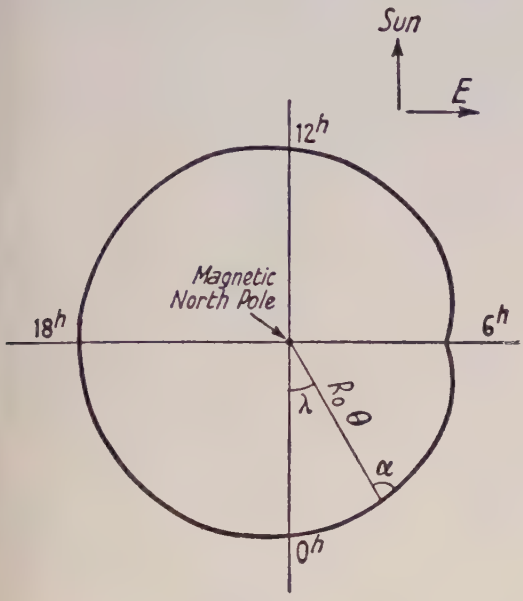


Fig. 2. Auroral curve (*I*-curve).

results are furtuitous or not. The answer depends upon how much the electric field is disturbed by the space charge, which is produced by the electrons and the ions when they enter the earth's magnetic field. Is it reasonable or not to approximate the resultant electric field as a homogeneous electric field?

It is very difficult to answer this question by a theoretical analysis. However, some light is thrown on the problem by an experiment by Malmfors (1946) which recently has been repeated by Block (1955). In some respects Malmfors-Block's experiment is similar to the famous experiment, which Birkeland made many years ago. Malmfors and Block placed a magnetized sphere ("terrella") in a chamber at low pressure and applied an electric field so as to reproduce the conditions assumed in the electric field theory. A beam of ionized gas was shot towards the terrella. It was found that eccentric luminous rings were produced around the poles as predicted by the theory. The rings are similar to the auroral zones. See Fig. 3.

It was found that even if no ionized gas was shot towards the terrella a similar phenomenon is obtained. In this case the ionization is produced by an electric discharge by the electric field. The discharge ionizes the gas mainly in the neighbourhood of the singular point x_d (in Fig. 1) and the electrons drift around the terrella following essentially a line which is similar to the border of the forbidden region. At the same time they oscillate along the lines of force and hit the terrella in the auroral zones.

In the terrella experiment the density of electrons and ions is certainly so large that space charge effects might be of importance. In spite of that the electrons move essentially according to the pattern, which has been derived without taking into account space charge effects. This indicates that space charge effects do not change the motion in the equatorial plane in a decisive way.

The reason for this may be that if space charge is accumulated in some region, it is discharged along the lines of force of the terrella. The electric field which is needed to send particles away along the lines of force is small compared to the field which is necessary in order to change the drift considerably.

Hence the terrella experiment indicates that we probably can treat the problem in the following way:

a. We calculate the motion in the equatorial plane under the influence of the given field E, taking no account of fields from space charge.

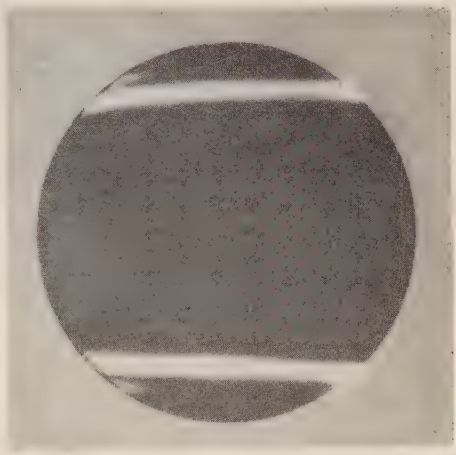


Fig. 3. Luminous rings on the terrella, seen from the 'nightside', Block, (loc. cit.).

b. The space charge, which is accumulated in certain regions, is assumed discharged along the lines of force to the earth (terrella).

To a second approximation we should also consider the electric fields from the space charge and the changes in the given magnetic field due to the currents.

§ 6. Motion of charged particles in a dipole field under the influence of an electric field

We study the motion of electrons and ions under the assumption that their motion is confined to the equatorial plane. The kinetic energy of the particles is supposed to be so small that the motion can be treated by the perturbation method. In the case of particles in the terrestrial field this means only that the energy must be below Cosmic Ray energies.

The motion of the particles consists of a circular motion which is superimposed upon a drift motion with the velocity

$$\vec{u} = -\frac{c}{eH^2}\vec{H}\times(\vec{f}+\vec{f^m}+\vec{f^i}) \tag{1}$$

with

$$\vec{f} = e\vec{E} \tag{2}$$

$$\vec{f^m} = -\mu \operatorname{grad} H \tag{3}$$

$$\vec{f^i} = -m\frac{\vec{du}}{dt} \tag{4}$$

$$\mu = \frac{W_\perp}{H} \tag{5}$$

where $W_\perp$ is the kinetic energy of the particle. During the motion μ remains constant. (See Cosmical Electrodynamics p. 13.)

In the earlier development of the theory (ALFVÉN 1939, 1940, 1950) the importance of the term f^i was not realized and this term was neglected. The inclusion of this term means that we take account of the inertia of the beam from the sun. We shall find that this gives rise to the initial phase of a storm.

Because of (4), equation (1) is a differential equation. In the cases of interest to us it is conveniently solved if we calculate a first approximation to u from (1) neglecting f^i, and later introduce the value of u into (4) and find a second approximation.

The electric field gives according to (1) and (2) a drift

$$u_x^E = 0 \tag{6}$$

$$u_y^E = -\frac{cE}{H} \tag{7}$$

The inhomogeneity of the magnetic field gives the drift

$$u_x^m = \frac{3ca\mu}{e}\cdot\frac{y}{HR^5} \tag{8}$$

$$u_y^m = -\frac{3ca\mu}{e}\cdot\frac{x}{HR^5} \tag{9}$$

Introducing as unit length

$$L = \left(\frac{a\mu}{|e|E}\right)^{1/4} \tag{10}$$

we obtain

$$u_x^m = \frac{3cE}{H}\left(\frac{L}{R}\right)^4\frac{y}{R}\cdot\frac{e}{|e|} \tag{11}$$

$$u_y^m = -\frac{3cE}{H}\left(\frac{L}{R}\right)^4\frac{x}{R}\frac{e}{|e|} \tag{12}$$

As we shall see in the following the inertia term (4) is of importance only in the region so far from the dipole that (7) is much larger than (8) and (9). Hence we may compute the inertia drift from the time derivative of (7):

$$\frac{du_y^E}{dt} = \frac{cE}{H^2}\cdot\frac{dH}{dy}\cdot u_y^E = \left(\frac{cE}{H}\right)^2\cdot\frac{3ay}{HR^5} \tag{13}$$

which gives

$$u_x^i = -\left(\frac{cE}{H}\right)^2\cdot\frac{3cma}{eH^2R^4}\cdot\frac{y}{R} \tag{14}$$

$$u_y^i = 0 \tag{15}$$

The total drift velocity u is composed of u^E, u^m and u^i

$$u_x = u_x^m + u_x^i \tag{16}$$

$$u_y = u_y^E + u_y^m \tag{17}$$

We introduce the distance

$$\lambda = (a/H_0)^{1/3} \tag{18}$$

where the dipole field equals the homogeneous field H_0 and put

$$\eta = \frac{R}{\lambda} \tag{19}$$

This gives

$$H = H_0(1 + \eta^{-3}) \tag{20}$$

and

$$u_x^i = -\frac{3mv_0^3}{eE\lambda} \cdot \frac{y}{R} \cdot \frac{\eta^8}{(1+\eta^3)^4} \tag{21}$$

with

$$v_0 = \frac{cE}{H_0} \tag{22}$$

which is the drift velocity in the beam far away from the dipole.

The value of u_x^i is small unless $\eta \approx 1$. When the particles drifting with the velocity (7) pass the region $\eta \approx 1$ they are displaced in the x-direction a distance x_λ which is:

$$x_\lambda = \int u_x^i dt = \int_\infty^0 \frac{u_x^i}{u_y^E} dy \tag{23}$$

or from (21) and (7)

$$x_\lambda = \frac{mv_0^2}{2eE} \int_\infty^0 \frac{6\eta^5}{(1+\eta^3)^3} \cdot \frac{y}{R} \frac{dy}{\lambda} \tag{24}$$

If x_λ is small so that x can be considered as approximately constant during the motion we have $ydy/R\lambda = d\eta$. The value of the integral is -1. This gives

$$x_\lambda = -\frac{mv_0^2}{2eE} \tag{25}$$

The displacement in the x-direction is so large that the increase eEx_λ in electrostatic energy exactly compensates the loss in kinetic energy $\frac{1}{2} mv_0^2$. As u_x^i decreases very rapidly on both sides of a maximum near $\eta \approx 1$ we could approximately describe what happens by saying that when the particles reach the distance λ from the dipole they are suddenly stopped and displaced x_λ.

§ 7. *Numerical values*

In order to estimate the numerical values of the important quantities, we tentatively put

$$v_0 = 2 \cdot 10^8 \text{ cm/sec} \tag{1}$$

a value which corresponds to a motion from the sun to the earth in $0.75 \cdot 10^5$ sec (≈ 1 day). Further we put

$$E = 1.4 \cdot 10^{-5} \text{ volt/cm} = 0.5 \cdot 10^{-7} \text{ e.s.u.} \tag{2}$$

a value which has been derived from the Cosmic Ray storm effect during a moderate storm (ALFVÉN, 1949, 1950, 1954). This gives from 6 (22)

$$H_0 = \frac{cE}{v_0} = 0.7 \cdot 10^{-5} \text{ gauss} \tag{3}$$

From 6 (18) we find with $a = 8 \cdot 10^{25}$ gauss cm^3:

$$\lambda = 2 \cdot 10^{10} \text{ cm} \tag{4}$$

The value of L could be derived from the polar distance Θ of the auroral zone. The polar distance of the auroral curve (compare Cosmical Electrodynamics, p. 182) varies between arcsin 0.872 $\sqrt{R_0/L}$ and arcsin 1.164 $\sqrt{R_0/L}$, so for a rough average Θ we have

$$L = R_0 \sin^{-2}\Theta \tag{5}$$

where R_0 is the earth's radius. With $\Theta \approx 22°$ this gives

$$L = 5 \cdot 10^9 \text{ cm} \tag{6}$$

Equations 6 (10), 6 (5), and 6 (18) give

$$\left(\frac{L}{\lambda}\right)^4 = \frac{W_\perp^0}{eE\lambda} \tag{7}$$

where $W_\perp^0$ is the kinetic energy of the circular motion of the electrons in the field H_0. This is a measure of the electron temperature in the beam. From (2) and (4) we find

$$eE\lambda = 2.8 \cdot 10^5 \text{ e volt} \tag{8}$$

The values (4) and (6) give $W_\perp^0 \approx 10^3$ e volt, corresponding to an electron temperature of 10^7 degrees. The ion temperature is probably much smaller.

The translational energy of the protons ($m = 1.6 \cdot 10^{-24}$g) in the beam is

$$\frac{m_H v_0^2}{2} \approx 2 \cdot 10^4 \text{ e volt} \tag{9}$$

and of the electrons

$$\frac{m_e v_0^2}{2} \approx 10 \text{ e volt} \tag{10}$$

From (2), (8), (9) and 6 (25) we find that for protons x_λ is of the order 10^9 cm, whereas for electrons x_λ is only 10^6 cm. Hence we can neglect the inertia drift of the electrons. We can also neglect quadratic terms of u^i/u^E for ions.

§ 8. Currents produced by the motion

The currents produced by the trochoidal motion of charged particles consist of the currents due to the spiralling motion and the currents due to the drifts. The magnetic field of the spiralling motion is the same as the field from the "equivalent dipoles". This corresponds to a diamagnetic effect of the gas and is usually not so important as the field produced by the drift currents so we shall primarily discuss the latter effect.

Suppose that per unit volume there are n^+ ions with charge e^+ and n^- electrons with charge e^-. The drift of the ions is u^+ and the drift of the electrons u^-. Then the current density is

$$j = n^+e^+u^+ + n^-e^-u^- \tag{1}$$

As according to § 6 u^+ and u^- have three components each it is convenient to split up the current into these three components:

$$j = j^E + j^i + j^m \tag{2}$$

As we shall see j^E and j^i are responsible for the initial phase and j^m for the main phase of a storm.

Putting $e = e^+ = -e^-$, we have from (1) and 6 (6)—6 (15)

$$\begin{cases} j_x^E = 0 & (3) \\ j_y^E = e(n^- - n^+)c\dfrac{E}{H} & (4) \end{cases}$$

$$\begin{cases} j_x^i = -(n^+m^+ + n^-m^-)\dfrac{3v_0^3}{E\lambda}\dfrac{\gamma}{R}\cdot\dfrac{\eta^8}{(1+\eta^3)^4} & (5) \\ j_y^i = 0 & (6) \end{cases}$$

$$\begin{cases} j_x^m = (n^+\mu^+ + n^-\mu^-)\,3ca\dfrac{\gamma}{HR^5} & (7) \\ j_y^m = -(n^+\mu^+ + n^-\mu^-)\,3ca\dfrac{x}{HR^5} & (8) \end{cases}$$

Here m^+ and m^-, μ^+ and μ^- refer to the ions and to the electrons.

Before the beam reaches the earth's magnetic field we must in average have $n^+ \approx n^- = n_0$ in order to avoid very large space charge.

During the motion in the earth's magnetic field the density will change. We can compute the density from the equation

$$\frac{\partial(nu_x)}{\partial x} + \frac{\partial(nu_y)}{\partial y} = 0 \tag{9}$$

As the inertia drift of the electrons is negligible, the electrons in the λ-region have only a velocity in the $-y$ direction. This gives

$$n^-u_y = n^-c\frac{E}{H} = \text{const.} \tag{10}$$

or

$$n^- = n_0\frac{H}{H_0} = n_0\frac{1+\eta^3}{\eta^3} \tag{11}$$

As the magnetic drift j^m goes in circles around the dipole, (11) is generally valid (except of course in the forbidden region).

For the ions the drift u_x^i gives the current (5) which is the x-component of a circular current the y-component of which is (4). This gives

$$j_x^i = -\frac{\gamma}{x}j_y^E \tag{12}$$

From (4) and (5) we obtain after simple calculations the ion density

$$n^+ = \frac{n}{1 + x_\lambda\dfrac{x}{R}\cdot P^i} \tag{13}$$

where

$$P^i = -\frac{1}{\lambda}\cdot\frac{6\eta^5}{(1+\eta^3)^3}$$

As $x_\lambda P^i \ll 1$ we get approximately

$$n^+ = n^-\left(1 - x_\lambda\frac{x}{R}P^i\right) = n_0\frac{1+\eta^3}{\eta^3}\left(1 - x_\lambda\frac{x}{R}P^i\right) \tag{14}$$

This satisfies (9) (neglecting quadratic terms).

According to (7) and (8), the magnetic drift gives also currents in circles around the dipole.

§ 9. The current system

Hence we obtain a system of *circular* currents:

$$i = j^i + j^m \tag{1}$$

Introducing 8 (11) we find from 8 (5), 6 (22) and 6 (20)

$$j^i = K^i P^i \tag{2}$$

with

$$K^i = \frac{c}{H_0} \cdot \frac{n_0 m_0 v_0^3}{2} \tag{3}$$

and

$$P^i = -\frac{1}{\lambda} \frac{6\eta^5}{(1+\eta^3)^3} \tag{4}$$

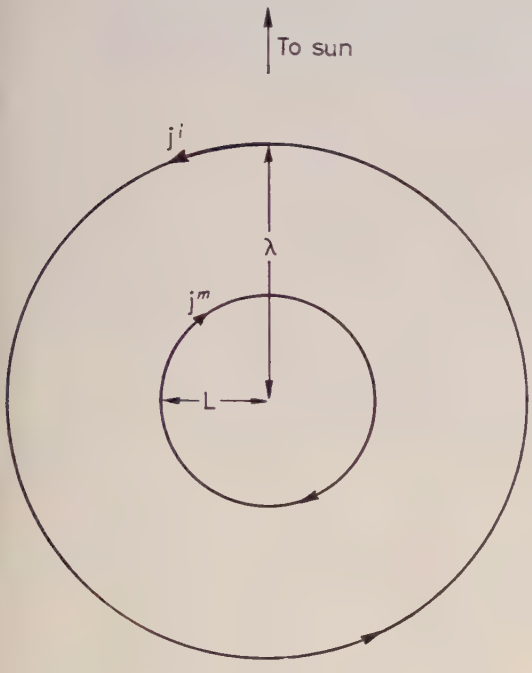


Fig. 4. Current system in the equatorial plane.

where $m_0 = m^+ + m^-$ is the mass of an ionized atom in the primary beam. Further we have from 8 (7), 8 (8), 8 (11), 6 (20) and 6 (22)

$$j^m = K^m P^m \tag{5}$$

with

$$K^m = e n_0 v_0 L_0 \tag{6}$$

and

$$P^m = \frac{3}{L_0} \left(\frac{L_0}{R}\right)^4 \tag{7}$$

where

$$L_0^4 = L_+^4 + L_-^4 = \frac{a}{eE}(\mu^+ + \mu^-) \tag{8}$$

As remarked earlier $\mu^+ \ll \mu^-$, which means that

$$L_0 \approx L_- = L$$

In a disk in the equatorial plane with the thickness 1 cm the total inertia current equals K^i and the magnetic drift current between $R = L_0$ and infinity equals K^m. The ratio between the two currents is

$$\vartheta = \frac{K^i}{K^m} = \frac{m_0 v_0^3}{2} \cdot \frac{1}{eEL} \tag{9}$$

The length L is related to the auroral zone and according to 7 (6) a reasonable value is $L = 5 \cdot 10^7$ cm. With 7 (2) and 7 (9) this gives

$$\vartheta = 0.3 \tag{10}$$

With this value given and $\lambda/L = 4$ it is possible to compute the current density as a function of the distance R from the dipole:

$$j = K^m(P^m + 0.3 P^i) \tag{11}$$

The result is plotted in Fig. 5. It is seen that outside $R \approx 0.75\ \lambda$ there is a current in the anti-clockwise direction tending to produce an increase in the magnetic field in the central parts of the equatorial plane. Inside $R \approx 0.75\ \lambda$ the current flows in the clockwise direction and gives a decrease in the magnetic field near the dipole. (Compare Fig. 4.)

According to (7) the current density goes towards infinity when we approach the dipole. However, the particles of the beam are deflected by the magnetic gradient drift so that there is a "forbidden region" around the dipole. The shape of the forbidden region for the electrons is shown by Fig. 1 which also shows the path of the electrons. The border of the forbidden region intersects the x-axis at the points $x = +\sqrt[4]{3}\, L = 1.32\, L$ and $x = -0.74\, L$ and the y-axis at the points $y = \pm 0.83\ L$. Inside the forbidden region the positive ions produce a current

$$\vec{j} = e n^+ (\vec{u_E^+} + \vec{u_m^+}) \tag{12}$$

For the positive ions there is a similar forbidden region which is the minor image of the forbidden region for the electrons, but enlarged or diminished in the proportion L^+/L^- where L^+ and L^- are given by 6 (10).

The development of a magnetic storm

§ 10. Deformation of the front surface.

We shall now discuss what happens when a beam approaches the dipole. We assume that the beam has a plane front surface so that

at a certain time t_0 there are no charged particles for $\gamma < \gamma_0'$ where γ_0' is supposed to be much larger than λ. For $\gamma > \gamma_0'$ the particle density is constant $n_0^+ = n_0^- = n_0$.

When $x = \infty$ the position γ_1 of the front at the time t is

$$\gamma_1 = \gamma_0 - v_0 (t - t_0') \tag{3}$$

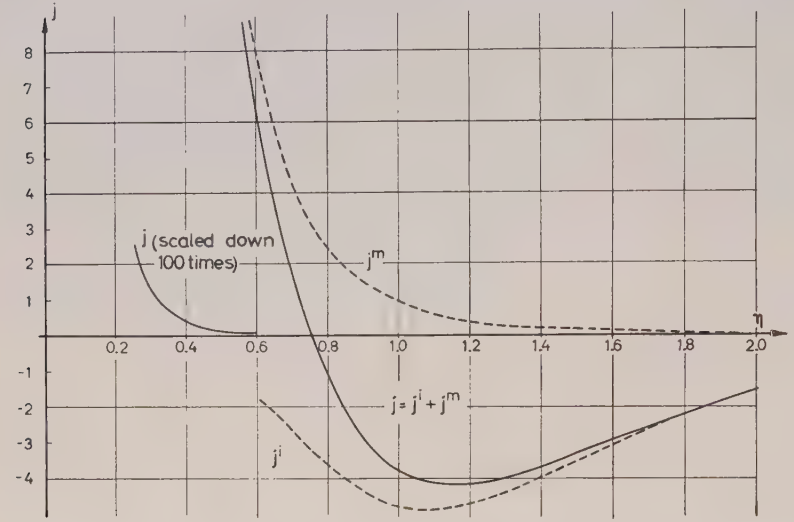


Fig. 5. The current density $j = j^i + j^m$ as a function of the distance $\eta = R/\lambda$ from the dipole.

We still neglect the electric fields produced by space charge. This means that our treatment holds only for the case when n_0 is infinitely small.

In a region where the magnetic drift is negligible (i.e. $R \gg L$) the front of the beam moves with the velocity

$$\frac{dy}{dt} = v^E = -\frac{cE}{H_0 + aR^{-3}} \tag{1}$$

Putting $\lambda = (a/H_0)^{1/3}$, $R = (x^2 + \gamma^2)^{1/2}$ and $v_0 = cE/H_0$ we obtain after integration

$$\gamma_0' - \gamma + \frac{\lambda^3}{x^2}\left(1 - \frac{\gamma}{R}\right) = v_0 (t - t_0') \tag{2}$$

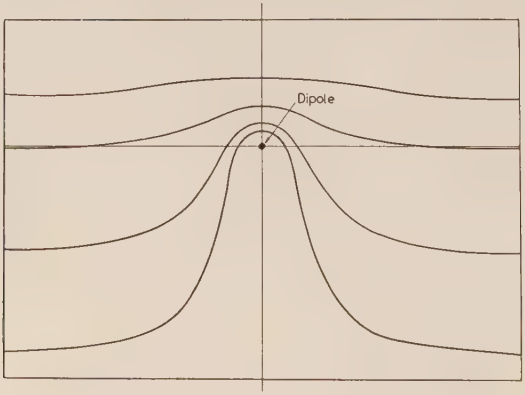


Fig. 6. Consecutive positions of the front surface of the beam in the equatorial plane.

The dipole field retards the front so that it lags behind a distance $\triangle\gamma$. If

$$\gamma = \gamma_1 + \triangle\gamma \tag{4}$$

we have

$$\triangle\gamma = \frac{\lambda^3}{x^2}\left(1 - \frac{\gamma}{R}\right)$$

or

$$\triangle\gamma = \frac{\lambda^3}{R^2 + R\gamma} \tag{5}$$

The shape of the front is given by (4) and (5). Fig. 6 shows its consecutive positions.

At the γ-axis the position of the front is given by

$$\gamma = \gamma_0' - v_0 (t - t_0') + \frac{\lambda^3}{2\gamma^2} \tag{6}$$

or

$$\frac{\lambda^3}{2\gamma^2} - \gamma = v_0 (t - t_0) \tag{7}$$

where t_0 is the time at which the point reaches $\gamma = \lambda/\sqrt[3]{2}$.

Thus the front forms a "hollow". Its geometry has some similarity with the hollow in Chapman-Ferraro's theory, but its electromagnetic properties are quite different. When the front has reached the distance L the magnetic drift becomes important. Electrons

moving anti-clockwise around the dipole will close the hollow and transform it into the forbidden region of Fig. 1.

§ 11. Current across the y—z plane

When the gas cloud approaches the dipole the currents near $R=\lambda$ will be developed first. This gives anti-clockwise currents, producing an increase in the magnetic field near the dipole. This phenomenon should be identified with the *initial phase*. When later the cloud has reached the vicinity of L clockwise currents are produced giving the decrease in the terrestrial magnetic field which is characteristic for the *main phase*.

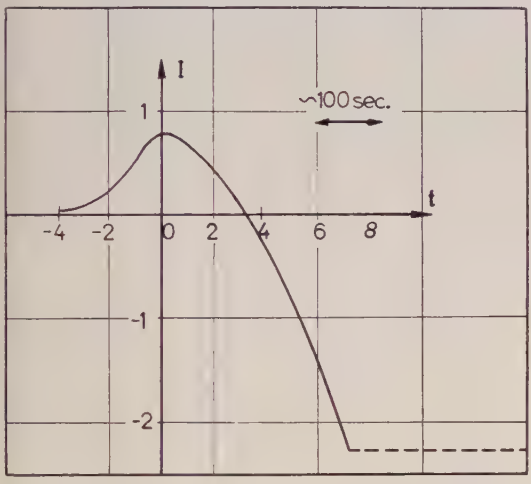


Fig. 7. The current across the y—z-plane as a function of time when the beam approaches the dipole.

In order to study the development of a storm it is important to calculate the electric charge transported across the y—z plane as a function of time. Hence we want to compute

$$I(t)=\int_{y}^{\infty} j\,dR \qquad (1)$$

where j is given by 9 (11) and y is a function of time according to 10(7).

As

$$\int_{y}^{\infty} P^i dy=\int_{\eta}^{\infty}\frac{6\eta^5 d\eta}{(1+\eta^3)^3}=\frac{1+2\eta^3}{(1+\eta^3)^2} \qquad (2)$$

and

$$\int_{y}^{\infty} P^m dy=\frac{3\lambda}{L}\int_{\eta}^{\infty}\left(\frac{L}{\lambda\eta}\right)^4 d\eta=\left(\frac{L}{\lambda}\right)^3\frac{1}{\eta^3} \qquad (3)$$

we find from (1)

$$I=K^m\left[\left(\frac{L}{\lambda}\right)^3\frac{1}{\eta^3}-\vartheta\frac{1+2\eta^3}{(1+\eta^3)^2}\right] \qquad (4)$$

which together with

$$\frac{1}{2\eta^2}-\eta=\frac{v_0}{\lambda}(t-t_0) \qquad (5)$$

defines the transport of charge across the y—z-plane as a function of time.

Fig. 7 shows I as a function of time. The current crossing the y-axis can be considered as consisting of a current at a distance λ, flowing in the $-x$ direction and a current at the distance L flowing in the x direction. The λ current starts at first and increases towards an asymptotic value. The L current starts some time later and soon causes the total current in the $-x$ direction to decrease to zero and then to a negative value. When the cloud has advanced to a distance of approximately L from the dipole the magnetic gradient drift deflects the particles. The conditions become more complicated. The main result is that the current does not increase any more.

When particles reach the negative y-axis behind the forbidden region an L current with the reverse sign begins to build up. Later also a λ current at $y=-\lambda$ starts. When the whole field (except the forbidden region) is filled up the total current across the y—z plane is zero and stationary conditions are reached. The λ current flows anti-clockwise along the whole circle $R=\lambda$ and produces an increase in H near the dipole which is more than compensated by the stronger clockwise current at the distance L.

The charge transported across the y—z plane by the λ current builds up a positive space charge at the circle $R=\lambda$ on the negative side of the y—z plane and a negative charge on the positive side. The λ current near $(x=-\lambda, y=0)$ and $(x=\lambda, y=0)$ is given by 10 (4), and the difference n^+-n^- is a result of the space charge. In the stationary state there is a space charge on the circle $R=\lambda$, which is constant in time. Its density is proportional to $-x$.

The space charge produces an electric field E' which inside the circle λ amplifies the primary field E. As we are still discussing the case when the density of the charged particles is so low that electric fields due to space charge

are negligible we have $E' \ll E$. Displacement currents due to the increase of E' close the current circuit until the stationary state is reached.

§ 12. *Discussion*

The theory which has been presented so far is a straight-forward calculation of what should happen when electrons and ions drift in the equatorial plane of a magnetic field 6 (1) under the influence of an electric field E. Up to now it has been assumed that the density is so low that electric fields from space charge and magnetic fields from currents are small compared to the given fields H and E.

In order to estimate the order of magnitude of the important quantities λ, L, ϑ, and the time-scale, it was necessary to introduce the values of E and H_0 of the beam. In order to avoid all *ad hoc* assumptions these values were taken from the theory of Cosmic Ray storm effects. The only value which was derived from magnetic storm observations and hence may by considered as an *ad hoc* assumption was the value of L which was derived from the polar distance of the auroral zone. (Also the assumption $\mu^+ \ll \mu^-$ or $L^+ \ll L^-$ has a similar character, but this assumption is not essential for the present discussion.)

The theory, which by neglecting space charge effects can at best be a first approximation, gives a current which at first flows anticlockwise and later is superimposed by a much stronger clockwise current. An identification with the initial phase and the main phase is natural.

The duration of the initial phase is according to Fig. 7 of the order of 10 minutes. This is less than one tenth of the observational value which is a few hours.

§ 13. *Production of space charge*

The current system which we have derived gives rise to space charge in some regions:

a. At the circle λ and around it positive charge is accumulated mainly on the $-x$ side of the y-axis and negative charge mainly on the positive side.
b. During the initial phase space charge is accumulated at or near the advancing front surface. In the λ region the space charge distribution of the stationary state is reached immediately when the front has passed. In the L region, however, a transient phenomenon occurs. When the L current across the y—z plane begins to flow the advancing front has the shape as shown in Fig. 6. The current flows along circles and produces a positive space charge at the $+x$ side of the advancing stream and a negative charge at the $-x$ side.
c. As the L system flows along circles but the forbidden region is limited by a non-circular curve, space charge is accumulated at the border of the forbidden region.

§ 14. *Currents along the lines of force*

The motion of charged particles which we have studied produces space charge in certain regions. When we go from the infinitely small densities which we have considered in § 6—12 to the densities which we have to deal with during a real storm, we find that the electric fields from the space charge become very large compared to E. There are two different possibilities:

a. The space charge effects disturb the whole type of motion so considerably that the accumulation of space charge is diminished very much.
b. The space charge produces currents along the lines of force and is discharged in this way without changing the motion in the equatorial plane.

The real case is probably somewhere between these extreme cases.

A detailed theoretical analysis of the problem is extremely difficult because it is likely that the discharge along the lines of force is associated with plasma oscillations, and at present our knowledge of these complicated phenomena is rudimentary. The assumption that plasma oscillations are important has been strengthened by Block's terrella experiment, in which the discharge easily produces "noise", showing the presence of plasma oscillations.

There is no doubt that the space charge will modify the motion of the particles to some extent. On the other hand the terrella experiment indicates that even when space charge is produced the change in the type of motion is not very large. This implies that most of the space charge is carried away along the lines of force.

Hence the picture we make of a magnetic storm is the following:

a. The motion of charged particles in the equatorial plane in the case of infinitely small density gives the basic pattern.
b. The space charge which is accumulated in some regions is discharged along the magnetic lines of force.
c. Some effects due to electric field from the space charge modify the motion but do not change it fundamentally.

If b) is applied to the currents in the L region we obtain the current system of the main phase. The circular currents in the equatorial plane close as currents along the lines of force to the auroral zones and as currents in the upper atmosphere in the auroral zones. This theory has been worked out long ago (ALFVÉN, 1939, 1940, 1950) and the magnetic field from this current system represents the real disturbances rather well over the whole earth. A possible exception is found inside the auroral zones where the real field possibly is somewhat stronger than the theoretical field (KIRKPATRICK, 1952). It is not quite clear whether this discrepancy is significant or due to some secondary effect which have not been considered. A possible explanation is that the discharge produced by the λ current gives an additional field.

If we apply the same principle to the λ current we find that the positive space charge accumulated at the negative side of the γ-axis and the negative charge at the positive side should discharge along the lines of force to the polar caps. They hit the ionosphere in a region which is the projection along the magnetic lines of force of the circle on the surface of the earth.

A current system of this type seems to give a magnetic disturbance which is reconcilable with the field measured during the initial phase.

§ 15. Space charge accumulated in the equatorial plane

According to our assumptions the space charge produced in the equatorial plane should be discharged along the lines of force to the polar caps. Hence the current flowing to one of the polar caps from the surface element $d\sigma$ of a layer of thickness $\triangle$ near the equatorial plane is

$$di = -\frac{\triangle}{2} \operatorname{div} j d\sigma \tag{1}$$

The current density is

$$j_z = \frac{di}{d\sigma} = -\frac{\triangle}{2} \operatorname{div} j \tag{2}$$

Here j means the total current density

$$j = j^E + j^i + j^m \tag{3}$$

As j^E is the same for electrons as for ions, it does not produce any space charge. Further j^m goes in circles around the dipole and does not either produce space charge (outside the forbidden region). Hence for the currents in the λ region we obtain from 8 (5), 8 (6) and 9 (2)

$$j_z = -\frac{\triangle}{2} \operatorname{div} j^i = -\frac{\triangle}{2} \cdot \frac{\partial j_x^i}{\partial x} = -\frac{\triangle}{2} K^i \cdot \gamma \frac{\partial}{\partial x}\left(\frac{P^i}{R}\right) \tag{4}$$

or as

$$\frac{\partial}{\partial x} = \frac{x}{\lambda^2 \eta} \frac{\partial}{d\eta} \tag{5}$$

we find

$$j_z = -\frac{\triangle}{\lambda^2} \cdot 3 K^i \sin \lambda' \cos \lambda' \eta \frac{d}{d\eta}\left[\frac{\eta^4}{(1+\eta^3)^3}\right] \tag{6}$$

or

$$j_z = \frac{\triangle}{\lambda^2} \cdot 3 K^i \sin \lambda' \cos \lambda' \phi \tag{6 a}$$

with

$$\phi = \frac{5\eta^3 - 4}{(1+\eta^3)^4} \eta^4 \tag{7}$$

Here λ' means the longitude angle counted positive eastwards from the midnight direction:

$$x = \lambda\eta \sin \lambda'; \quad y = -\lambda\eta \cos \lambda' \tag{8}$$

The function $\phi(\eta)$ is plotted in Fig. 8. It is evident, that in the first and third quadrants a large positive space charge is accumulated outside $\eta = \sqrt[3]{0.8}$, whereas there is a small negative charge inside this limit. In the second and fourth quadrants the signs are reversed.

The physically important phenomenon is that the circular λ current of § 9

$$j^i = K^i P^i \tag{9}$$

consists of one component

$$j_1^i = K^i P^i \cos^2 \lambda' \tag{10}$$

which is produced by the relative motion of the ions and the electrons and another component

$$j_2^i = K^i P^i \sin^2 \lambda' \qquad (11)$$

which is produced by the displacement of a net space charge. As at the densities we have to consider in reality the latter effect becomes negligible, only the first component is of importance. This current is closed by currents along the lines of force with the current density

$$j'_z = \frac{\triangle}{2} K^i P^i \frac{d \cos^2 \lambda'}{\lambda \eta d\lambda'} \qquad (12)$$

$$j'_z = \frac{\triangle}{\lambda^2} K^i \sin \lambda' \cos \lambda' \frac{6\eta^4}{(1+\eta^3)^3} \qquad (13)$$

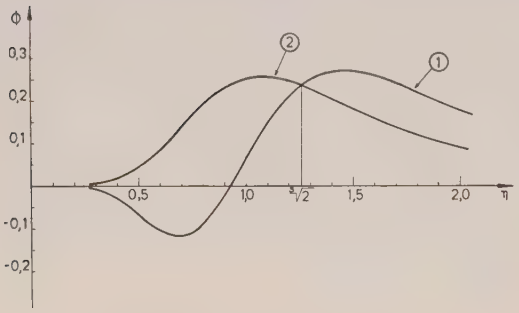


Fig. 8. Space charge accumulated in the equatorial plane.
Curve 1 calculated from eq. 15 (7)
Curve 2 " " " 15 (13)

Superimposed on this effect is a relatively small effect due to the radial component of j_2^i, which carries charge outwards. This effect changes (13) into (6). It is not so sure that this effect is more important than other secondary effects which we have neglected. For example, if the emission of space charge from the equatorial plane is delayed somewhat, the effect may vanish and (13) may be closer to reality than (6). Equation (13) is also represented in Fig. 8.

§ 16. Projection of the λ region upon the earth's surface

In order to find how the λ current closes over the polar cap we must calculate the polar distance Θ where the earth's surface is reached by a line of force which passes the equatorial plane at the distance $\eta\lambda$.

The flux ϕ_1 which penetrates the earth's surface between the parallel circle Θ and the equator is

$$\phi_1 = \left(\frac{2a}{R_0^3} - H_0\right)(\pi R_0^2 - \pi R_0^2 \sin^2 \Theta) \qquad (1)$$

The flux ϕ_2 through the equatorial plane between the circle $\eta\lambda$ and the circle R_0 is

$$\phi_2 = \int_{R_0}^{\eta\lambda} \left(H_0 + \frac{a}{R^3}\right) 2\pi R dR = \frac{2\pi a}{R_0} - \frac{2\pi a}{\eta\lambda} + \pi(\eta^2\lambda^2 - R_0^2) H_0 \qquad (2)$$

The condition $\phi_1 = \phi_2$ gives

$$\sin^2 \Theta = \frac{\eta_0}{\eta} \cdot \frac{2 - \eta^3}{2 - \eta_0^3} \qquad (3)$$

with

$$\eta_0 = \frac{R_0}{\lambda} \qquad (4)$$

$$\lambda = \left(\frac{a}{H_0}\right)^{1/3} \qquad (5)$$

As $\eta_0 \ll 1$ we have approximately

$$\sin^2 \Theta = \frac{\eta_0}{\eta}\left(1 - \frac{\eta^3}{2}\right) \qquad (6)$$

This shows that in the equatorial plane the circle

$$\eta = \sqrt[3]{2} \qquad (7)$$

is the border between those lines of force which go to the earth and those going to infinity. Only space charge produced inside (7) can be discharged over the polar cap. The lines of force through this circle go to the pole.

Differentiating (6) we obtain

$$\frac{d\eta}{d\Theta} = -\frac{\eta^2 \sin 2\Theta}{\eta_0(1+\eta^3)} \qquad (8)$$

A surface element $d\sigma$ in the equatorial plane is

$$d\sigma = \lambda^2 \eta d\eta d\lambda' \qquad (9)$$

where λ' is the longitude angle. A surface element of the polar cap is

$$dS = R_0^2 \sin \Theta d\Theta d\lambda' \qquad (10)$$

where R_0 is the earth's radius. If a current di along the lines of force has the current density

$$j_z = \frac{di}{d\sigma} \qquad (11)$$

when it leaves the equatorial plane, its density is J when it reaches the polar cap:

$$J = \frac{di}{dS} \tag{12}$$

The relation between j and J is

$$JR_0^2 \sin\Theta d\Theta = j_z \lambda^2 \eta d\eta \tag{13}$$

or

$$J = \frac{2\eta^3 \cos\Theta}{\eta_0^3 (1+\eta^3)} j_z \tag{14}$$

§ 17. Currents to the polar caps

Introducing 15 (6) and 15 (13) into 16 (14) we obtain

$$J = \frac{6K^i \cos\Theta}{\eta_0^3} \sin\lambda' \cos\lambda' \cdot \Psi \tag{1}$$

with

$$\Psi = \frac{(5\eta^3 - 4)\eta^7}{(1+\eta^3)^5} \tag{2}$$

if we derive it from 15 (6)
or

$$\Psi = \frac{2\eta^7}{(1+\eta^3)^4} \tag{3}$$

if derived from 15 (13). These equations should be combined with 16 (3) in order to give the current as a function of the polar distance Θ. The formula (2) is derived if we apply our assumptions in a straight-forward way. However, it means that along the lines of force from the λ-region there should flow currents with opposite signs very close together. This is perhaps not very likely. The charges will probably get mixed so that what reaches the polar cap is more like (3) than (2). Both functions are plotted in Fig. 9, assuming

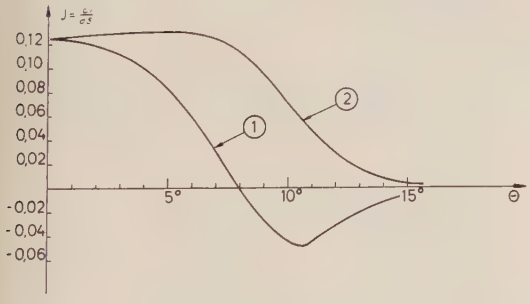


Fig. 9. Current density to the polar cap.
Curve 1 calculated from eq. 17 (2)
Curve 2 " " " 17 (3)

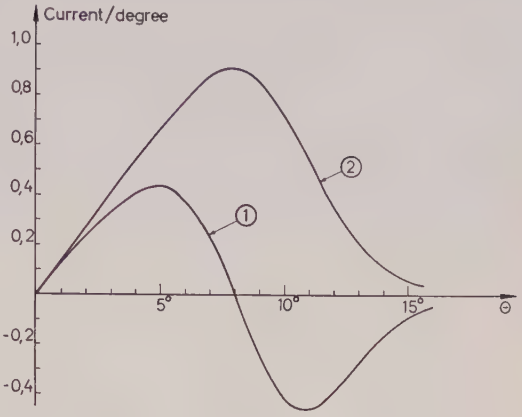


Fig. 10. Current/latitude degree to the polar cap.
Curve 1 calculated from eq. 17 (2) multiplied by $\sin\Theta\cos\Theta$
Curve 2 calculated from eq. 17 (3) multiplied by $\sin\Theta\cos\Theta$

$$\eta_0 = \frac{R_0}{\lambda} = 0.03$$

Fig. 10 shows current per latitude degree to the polar cap

$$\frac{di}{d\Theta} \sim \Psi \sin\Theta \cos\Theta$$

It should be observed that only the charge to the left of $\eta = \sqrt[3]{2} = 1.26$ in Fig. 8 is discharged over the polar cap. The rest should move away along the lines of force of the interplanetary field. Also in this respect a certain mixing may take the place, so that eventually even some of the charge outside $\eta = \sqrt[3]{2}$ is discharged over the polar caps.

The currents have maxima not very far from $\eta = 1$ which projected on the polar cap by means of 16 (3) gives

$$\sin^2\Theta \approx \frac{\eta_0}{2} \tag{4}$$

or from 16 (4) and 16 (5)

$$\Theta \approx \left(\frac{1}{2} R_0\right)^{1/2} \left(\frac{H_0}{a}\right)^{1/6} \tag{5}$$

Introducing the earth's magnetic field at the equator

$$H_e = \frac{a}{R_0^3} \tag{6}$$

we find

$$\Theta = 0.7 \left(\frac{H_0}{H_e}\right)^{1/6} \qquad (7)$$

Hence the polar distance is determined by the strength of the magnetic field H_0 in space outside the earth.

Introducing the value of H_0 from 7 (3) and putting $H_e = 0.3$ gauss we find

$$\Theta = 0.7 \cdot \left(\frac{7 \cdot 10^{-6}}{0.3}\right)^{1/6} = 0.12 = 7^\circ \qquad (8)$$

§ 18. On the existence of an "inner auroral zone"

According to the electric field theory the main phase of a magnetic storm is mainly due to a ring current in the equatorial plane at the distance $L \approx 5 \cdot 10^9$ cm which is closed by currents along the lines of force to the auroral zones at $\Theta \approx 20^\circ$. As shown earlier (ALFVÉN, 1939, 1940, 1950) the magnetic field of this current system gives a good agreement with the magnetic disturbances which are actually observed and it also accounts for some important features of the aurorae.

In the present paper we have found that besides the L current system we should expect another system at a distance of about $\lambda \approx 2 \cdot 10^{10}$ cm. This phenomenon is a direct effect of the inertia of the storm-producing beam. The currents in the equatorial plane are likely to close in a way similar to the L current. If this occurs we should expect *aurora at a polar distance of 5°—10°*, corresponding to the projection of λ on the earth's surface.

Hence besides the ordinary auroral zone at a polar distance of $\Theta \approx 20^\circ$ we should expect an *inner auroral zone* with a polar distance which is less than half of the ordinary auroral zone. According to 17 (7) its polar distance is determined by the interplanetary field H_0.

Aurorae in the inner auroral zone should start simultaneously with the initial phase of a storm, whereas aurorae in the ordinary auroral zone should essentially be associated with the main phase. Even during the main phase there should be aurorae in the inner zone. The auroral activity should be accompanied by geomagnetic disturbances, which should have a secondary maximum in the inner auroral zone. As according to § 9 the total current in the λ-region is a few tenths of the current in the L-region, we should expect the auroral activity and the magnetic activity to be a few tenths of those of the ordinary auroral zone.

These theoretical predictions should be checked by observations. Present observational data seem to give some support of the theory but this is far from conclusive. Certainly aurorae are frequently observed inside the ordinary auroral zone, but it is not known whether the frequency decreases in a monotonous way towards the pole or—as predicted by the theory—has a secondary maximum at 5°—10°. SUCKSDORFF (1947) has analyzed the geomagnetic results from some polar stations and concludes that there are very strong currents near the geomagnetic pole. VESTINE (1947) has reported one case when there is a strong magnetic disturbance at Thule and nothing at other stations. This gives additional support for the existence of an inner auroral zone. Of course much more material is necessary before it is legitimate to draw any certain conclusions.

REFERENCES

ALFVÉN, H., 1939: A Theory of Magnetic Storms and Aurorae, I, *Kungl. Sv. Vetenskapsakad. Handl.* (3), **18**, No. 3.

— 1940: A Theory of Magnetic Storms and Aurorae, II, III, ibid. **18**, No. 9.

— 1949: On the Solar Origin of Cosmic Radiation, *Phys. Rev.*, **75**, No. 11.

— 1950: On the Solar Origin of Cosmic Radiation, II, ibid. **77**, No. 3.

— 1950: Cosmical Electrodynamics, *Oxford.*

— 1954: On the Origin of Cosmic Radiation, *Tellus*, **6**, No. 3, 232.

BLOCK, L., 1955: Model Experiments on Aurorae and Magnetic Storms, *Tellus*, **7**, 65.

BRUNBERG, E. Å. and DATTNER, A., 1954: Variations of the Cosmic Ray Intensity during a magnetic storm, *Tellus*, **6**, 254.

KIRKPATRICK, C. B., 1952: On Current Systems Proposed for S_D in the Theory of Magnetic Storms, *Journal Geoph. Res.*, **57**, No. 4.

MALMFORS, K. G., 1946: Experiments on the Aurorae, *Ark. f. mat., astr. o. fysik*, **34B**, No. 1.

SUCKSDORFF, E., 1947: Diurnal Variations of Computed Electric Currents in the High Atmosphere, *Terr. Magn. and Atmosph. El.*, **52**, No. 2.

VESTINE, E. H., LANGE, I., LAPORTE, L. and SCOTT, W. E., 1947: The Geomagnetic Field, its Description and Analysis, *Carn. Inst. Publ.*, No. 580, Washington D. C

Model Experiments on Aurorae and Magnetic Storms

by L. BLOCK, The Royal Institute of Technology, Stockholm

(Manuscript received November 4, 1954)

Abstract

Model experiments on Alfvén's electric field theory of aurorae and magnetic storms have been performed. The transformation laws from nature to model are considered. It is concluded, that a proper scaling down of every detail in nature is not possible. Only the most significant features are scaled down properly. It is shown, that a mechanism like Alfvén's proposed auroral mechanism is very likely to be present in the model. In particular the conception seems to be experimentally confirmed, that space charge accumulated in the discharge region rather than destroy the whole pattern of motion, discharges along the magnetic field lines.

Introduction

In Alfvén's electric field theory of aurorae and magnetic storms (ALFVÉN, 1939, 1940, 1950) it is assumed, that a gaseous discharge due to an interplanetary electric field takes place in the surroundings of the earth during a storm. This discharge is affected by the terrestrial magnetic field and reaches the earth's atmosphere only in the auroral zones. The currents of the discharge are responsible for the magnetic disturbances observed during a storm.

Previously MALMFORS (1946) has made some experiments on this theory. In a cylindrical vacuum chamber (diam. 40 cm, length 75 cm), covered at the ends with glass plates, he arranged a magnetized steel sphere (diam. 10 cm) in the centre of the chamber and two insulated metal plates (60 × 25 cm) vertically on each side of the sphere separated by 24 cm. He found that a discharge between the two plates had many similarities with the discharge considered in the theory. Indeed, the results seemed so promising, that it was thought worth while to make a more detailed experimental study on a somewhat larger scale.

Experimental arrangement

A total view of the apparatus used is seen in fig. 1. The volume of the vacuum chamber is approximately 1.5 m^3. A pressure of about

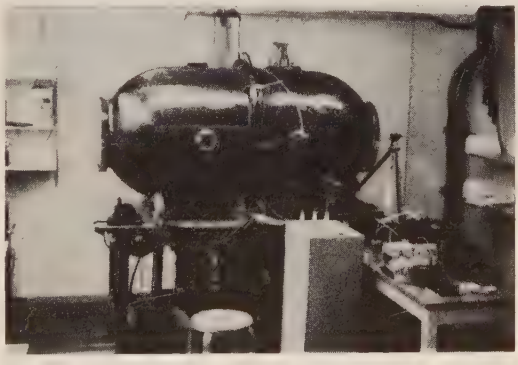


Fig. 1. Total view of the apparatus.

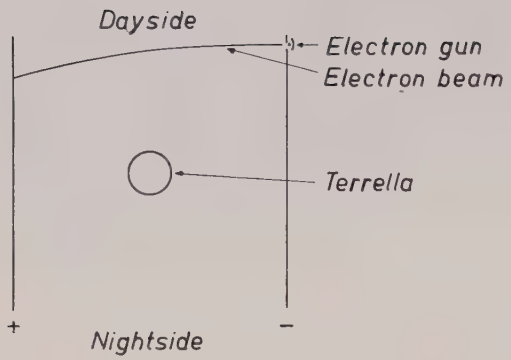


Fig. 2. Terrella and electrodes seen from above. The dipole direction of the terrella is perpendicular to the paper with the magnetic south pole above and the north pole below the paper.

10^{-4} mm Hg could be obtained in it at best. Fig. 2 is a sketch of the electrodes and the terrella, seen from above. They are arranged in exactly the same way as in Malmfors' experiment except that the dimensions are larger here. The electrodes are aluminum plates (60 × 60 cm). Only the terrella is of the same size as that of Malmfors, the diameter being 10.6 cm, but here a coil is used inside the terrella to produce the magnetic field, instead of Malmfors' permanently magnetized steel sphere. Thus we can vary the field and make it stronger than in Malmfors' experiments. The coil is so designed, that it can produce a rather good dipole field. The deviation from the exact dipole field is less than 5 % everywhere outside the terrella. Cooling by air is provided, because when 3,000 gauss at the poles is produced, the heating by the current amounts to 8 kW. An automatic switch is used to break the current, when the temperature of the windings reaches a certain value (150—200° C) in order to avoid damage. The temperature is determined by measuring the coil resistance.

Various probes and other electrodes for special investigations can if needed be placed in the discharge region.

Transformation from nature to model

Before describing the various experiments made in the model we must consider in some detail, how such experiments should be carried out in order to simulate a 'real' discharge of terrestrial dimensions in a proper manner. If the linear dimensions of a gaseous discharge are changed, certain similarity laws are valid (see COBINE, 1941, p. 209, ENGEL and STEENBECK, 1932, p. 95, or ALFVÉN, 1950, p. 39). In nature the following data are assumed to be valid (ALFVÉN, 1955).

Earth's diameter ..... $D = 1.3 \cdot 10^7$ m,
velocity of the ionized beam from the sun .. $v = 2 \cdot 10^6$ m/s,
sun's magnetic field at the earth's orbit ..... $B_0 = 0.7 \cdot 10^{-5}$ gauss,
electric field of the discharge ..... $E = 1.4 \cdot 10^{-3}$ V/m,
earth's magnetic field at poles ..... $B_p = 0.6$ gauss.

According to the similarity laws all potentials should be unaltered, because ionization and excitation potentials are the same. Therefore, all electric fields are inversely proportional to the linear dimensions, and according to Maxwell's equations, the same is valid for the magnetic fields. Now, in the actual case, the linear dimensions are diminished by a factor $k = 10^8$. Hence, the magnetic field strength at the poles of the terrella ought to be of the order of 60 million gauss. In nature the potential over a length of the order of the earth's diameter is $V_D = D \cdot E = 18$ kV. The normal distance between the electrodes of the model discharge is 5.4 terrella diameters, and so the applied voltage should be of the order of 100 kV.

Since these strong magnetic fields are impossible to obtain, we will neglect the condition about the unchanged potentials and put the question: How are the magnetic and electric fields connected, if the electron and ion orbits are unaltered, assuming them to be undisturbed by collision processes?

First we note that we can confine ourselves to the non-relativistic case, because the energies of the particles are at most 100 keV.

An electron moving in a magnetic field B with the velocity v perpendicular to the magnetic field lines has an orbital radius of curvature

$$\varrho = \frac{mv}{eB} \qquad (1)$$

Rationalized Giorgi-units are used throughout. If the kinetic energy $1/2\ mv^2$ is assumed to be proportional to the electric field E or the applied discharge voltage V_d, we find that if

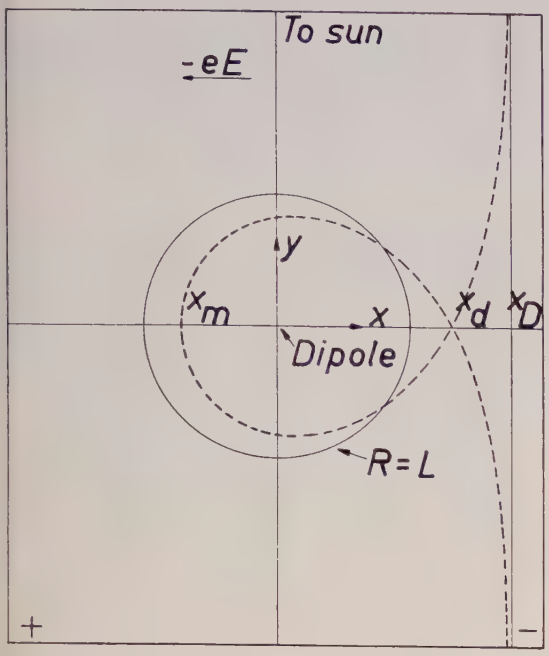


Fig. 3. Border line of the forbidden zone in the equatorial plane. Compare fig. 6.1, p. 180 in Cosmical Electrodynamics.

$x_D = 1.76$ L, $x_d = 1.32$ L, $x_m = -0.74$ L.
$x_D - x_d = 0.44$ L, $x_D - x_m = 2.50$ L.

$$\frac{V_d}{B^2} = \text{constant} \qquad (2)$$

the electron orbits remain unchanged. This will be illustrated with some examples, which are closely connected with the theory of aurorae and magnetic storms.

In a gaseous discharge two classes of electrons are usually present. Some electrons start from the cathode and acquire a high energy in the cathode fall. When they enter the plasma, their energy is of the same order as the whole discharge potential. We will call these electrons primary electrons. If the mean free path in the plasma is of the same order as the distance between cathode and anode, most of the primary electrons will reach the anode with only slightly diminished energy. A strong magnetic field perpendicular to the electric field may, however, prevent the primary electrons from penetrating deep into the plasma without a considerable loss of energy. The other class of electrons consists of secondary electrons formed in the plasma due to ionization by primaries, other secondaries or photons.

The primary electrons can then be considered as entering the magnetic field at the boundary between cathode fall and plasma with an initial energy equal to the cathode fall or in most cases approximately the discharge voltage. We see at once that (2) is applicable in this case. The secondaries on the other hand start in an electric field E and a magnetic field B, and we will assume their initial energy to be zero. Such an electron then starts moving in a trochoidal orbit with the mean energy perpendicular to the magnetic field expressed in electron volts

$$V'_\perp = \frac{m}{e}\left(\frac{E_\perp}{B}\right)^2 = 5.7 \cdot 10^{-12}\left(\frac{E_\perp}{B}\right)^2$$

where $E_\perp$ is the electric field component perpendicular to the magnetic field.

If we combine (1) and (3) we see that because $v \sim \sqrt{V'_\perp}$, we have

$$\varrho_{\text{sec}} \sim \frac{\sqrt{V'_\perp}}{B} \sim \frac{E_\perp}{B^2} \sim \frac{V_d}{B^2} \qquad (4)$$

and (2) is applicable also in this case provided the initial energy is small compared with $eV_\perp$ as calculated from (3).

Further the so called Störmer-unit of length (Störmer, 1933) is unchanged if the particle energy eV is proportional to B^2. Störmer has defined a characteristic length for the particle orbits in a magnetic dipole field

$$C = k\sqrt{\frac{a}{B \cdot \varrho}} \qquad (5)$$

where a is the magnetic dipole moment, $B\varrho$ is characteristic of the particle and k is a constant, which depends on the unit system. Now $B\varrho$ is proportional to the velocity of the particle and $B \sim a$, so that

$$C \sim \left(\frac{B}{V^{1/2}}\right)^{1/2} \qquad (6)$$

Alfvén has calculated the electron orbits in the equatorial plane of the dipole field. He has introduced a characteristic length (see e.g. Cosmical Electrodynamics p. 179)

$$L = \sqrt[4]{\frac{\mu_0}{4\pi} \cdot \frac{\mu a}{eE}} \qquad (7)$$

where a is the magnetic dipole moment, E is the electric field, $\mu_0 = 4\pi \cdot 10^{-7}$ henry/m is the permeability of vacuum and μ (not to be confused with μ_0) is the orbital magnetic moment of the particle

$$\mu = \frac{eV_\perp}{B} \tag{8}$$

which can be shown to be a constant of the motion. Here $eV_\perp$ is the kinetic energy of the particle referred to a system where the centre of curvature is at rest, i.e. where the drift velocity is zero. $V_\perp$ must not be confused with $V'_\perp$ of eq. (3). In the special case of eq. (3). $V_\perp = V'_\perp/2$. In most other cases the difference between $V_\perp$ and $V'_\perp$ is very small, although always $V_\perp \leq V'_\perp$. From equation (3) we see that

$$\mu \sim \frac{E^2}{B^3} \sim \frac{V_d^2}{B^3} \tag{9}$$

Thus we find

$$L^4 \sim \frac{V_d^2}{B^3} \cdot \frac{B}{V_d} = \frac{V_d}{B^2} \tag{10}$$

According to Malmfors (see MALMFORS, 1946 equation (7) or appendix II.) the latitude angle of the auroral zones should be a function of E/B^2.

Thus we see that we always arrive at the same connection between the applied voltage of the discharge and the magnetic field of the dipole

$$V_d = kB^2 \tag{11}$$

Particle energies in nature

The particle energies in the beam from the sun determine to a considerable extent the properties of the discharge in terrestrial regions. Therefore, we will consider these energies in this section. These considerations are of course very uncertain. Their purpose is only to permit us some familiarity with the problems of the discharge, rather than to obtain any definite results.

From the time difference between a solar flare and the commencement of a magnetic storm, the speed of the beam front from the sun can be calculated as $10^6 - 2 \cdot 10^6$ m/s. Now $2 \cdot 10^6$ m/s means an electron energy of 11 eV and a proton energy of 20 keV associated with the translation of the beam. These are lower limits, because the particles may travel in curved paths, e.g. trochoids. It seems very improbable, that the particles should go in approximately straight lines from sun to earth. It is worth while to note, that from the relation

$$eV = \frac{3}{2}kT, \tag{12}$$

one finds that 7,700° K is equivalent to 1 eV. This means that if the beam is ejected from the corona, where the temperature may be of the order of 10^6 °K, the electronic energies may be some 100 eV, which is significantly greater than the above mentioned value of 11 eV.

From the penetration depth of the auroral particles in the atmosphere Störmer and others have estimated the energy to 50—100 keV. Measurements made by MEINEL (1951) on the H_α-line found in auroral archs indicate that the impact velocity of the protons seems to be

$$v_H \geq 3{,}300 \text{ km/s}$$

or

$$W_H \geq 57 \text{ keV},$$

and the velocities perpendicular to the impact direction seem to be distributed between 0 and 1,000 km/s.

According to Alfvén's theory the particles penetrating the earth's upper atmosphere are accelerated in the electric field of the beam. If the border of the forbidden zone is 7 earth radii from the earth, the electric field must be of the order of 10 kV per earth radius or 10^{-3} V/m, which agrees well with the values calculated earlier in this paper.

From equation (7) we see that the currents of the discharge are determined to a great extent by the electron energy perpendicular to the magnetic field in the equatorial plane. We rewrite (7) in a somewhat different form

$$L^4 = \frac{V_\perp}{B} \cdot \frac{B_{eq} \cdot R^3}{E} \tag{7a}$$

$eV_\perp$ = electron energy perpendicular to the magnetic field, referred to a zero drift velocity system.
B_{eq} = earth's magnetic field at equator on surface,
R = earth's radius,
E = electric field of the beam.

Further we may note, that $V_\perp/B$ is constant along the path of the electron.

We will now find a reasonable value for $V_\perp$

at the border of the forbidden region in the equatorial plane. This has approximately the distance L from the dipole. Thus we have

$$L^4 = \frac{V_\perp \cdot B_{eq} \cdot R^3}{\frac{R^3}{L^3} B_{eq} E} = \frac{V_\perp}{E} L^3$$

or

$$L \cdot E = V_\perp \qquad (7\,b)$$

at the distance L from the earth's dipole. This occurs at two points on the border of the forbidden region. In fact the energy $eV_\perp$ of the electrons varies along the border line according to the relation

$$V_\perp = E(x_D - x) \qquad (7\,c)$$

See fig. 3. Hence we find that

$$2.5\,LE > V_\perp > 0.44\,LE \qquad (7\,d)$$

This means that, according to the values given above, $V_\perp$ may be of the order of 100 kV. If L is assumed to be seven earth radii, the magnetic field at the border line is approximately 10^{-3} gauss, and as $V_\perp/B$ is constant, we find that at a magnetic field of 10^{-5} gauss, i.e. in the beam outside the earth's magnetic field, $V_\perp = 1{,}000$ volts.

According to these strictly mathematical considerations these high energy electrons would have nearly all their kinetic energy perpendicular to the magnetic field. If we assume that μ = constant of the motion, also when the particles move along the field lines from the equatorial plane towards the auroral zones, the diamagnetic repulsion of the terrestrial magnetic field would be very strong (proportional to $V_\perp$) and prevent the particles from reaching the auroral zones. Therefore, one must assume that particles of all energies are present in the beam with a mean energy, which determines the dimensions of the forbidden zone and hence the auroral zones. It must also be assumed, that a statistical distribution of the velocities is maintained, mainly by plasma oscillations, everywhere in the discharge in spite of the strongly increasing $V_\perp$ in the earth's magnetic field. If the velocity distribution of the particles is isotropic, we are led to another constant of the motion than the μ of equation (8) as is shown in appendix I

$$\varkappa = \frac{V^{3/2}}{B} \qquad (13)$$

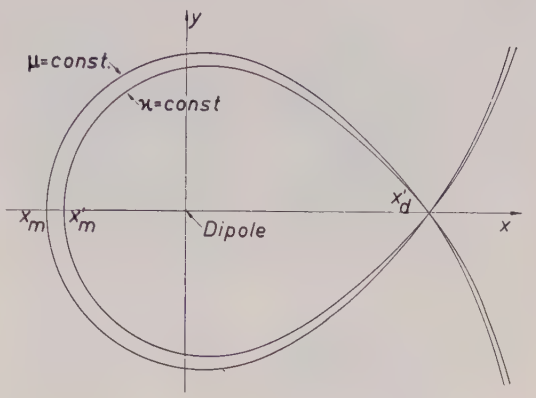


Fig. 4. Comparison between forbidden zones

a) assuming the orbital magnetic moment of the particles to be constant,

b) assuming isotropic velocity distribution everywhere $x'_d = 1.26\ L$, $x'_m = -0.63\ L$.

Here eV is the total energy (referred to a zero drift velocity system) and is therefore according to our assumptions equal to $3/2\ eV_\perp$. It is also shown in the appendix, that this leads to another form of the forbidden zone

$$x_0 - x = \frac{L^3}{R^2}$$
$$L^3 = \frac{1}{E}\left(\frac{\mu_0}{4\pi} a\varkappa\right)^{2/3} \qquad (14)$$

instead of

$$x_0 - x = \frac{L^4}{R^3}$$
$$L^4 = \frac{\mu_0}{4\pi} \cdot \frac{a\mu}{eE} \qquad (15)$$

for μ = constant. The forbidden zones in the two cases are compared in fig. 4. The difference is very small, the eccentricity being a little smaller for μ = constant.

Transformation of pressure from nature to model

As will be seen in the next section the pressure used in the experiments was about 10^{-3} mm Hg. This corresponds to a particle density of about 10^{13} particles/cm^3, and according to the transformation laws, the particle density should be inversely proportional to the linear dimensions. Since the scale factor is 10^8, the above mentioned particle

density in the experiments corresponds to 10^5 particles/cm^3 in nature. In reality the discharge takes place in regions with strongly varying pressures, the extremes of which are the ionosphere with 10^{12} or 10^{13} particles/cm^3 in the *E*-layer and the interplanetary space, where the density is probably much lower than 10^5 particles/cm^3. Therefore, the pressure used may be a good compromise between the two extremes.

Performed experiments

The various experiments performed will now be described, and after that we can apply the above considerations to the results obtained.

There are two kinds of discharges which have been studied. If the pressure was low ($<10^{-3}$ mm) there was an invisible non-self-sustained discharge (see COBINE, 1941, Ch. VII), which could only be maintained by means of some ionizing device, such as an electron beam indicated in fig. 2. The mean free path of the high energy electrons was then of the same order as the distance between anode and cathode. If the pressure was higher ($>2 \cdot 10^{-3}$ mm) it was possible to get a visible glow discharge without any extra ionizing device. If the visible discharge was used, the cathode voltage V_c was of the order of -2 kV with respect to the terrella (every voltage is given with respect to the terrella if not otherwise specified), the anode voltage $+20$ to 50 volts (it could not be higher because then the anode currents drawn from the plasma would be too high), and the currents were about 10—50 mA. In the case of the invisible discharge, however, V_c was usually -10 kV, $V_a = 3$—5 kV and the currents only 0.1—1 mA. It is probable, that the invisible discharge has more to do with the phenomena in nature than the visible one. This question will be considered later.

In figs. 5—11 photographs of the various discharges are shown. Fig. 5 shows the visible discharge with increasing magnetic field *B*. At zero magnetic field the electrons move in straight lines from the cathode to the terrella, which is covered by fluorescent material. Thus only the terrella side facing the cathode is bright. If a weak magnetic field (20 gauss at poles) is turned on, the electrons are deflected through a small angle, and the bright hemisphere of the terrella is turned over through the same angle to the dayside. With increasing magnetic field we get more complicated figures. At about 300 gauss, however, the equator becomes part of a forbidden zone, and 'auroral zones' begin to develop. With *B* still more increasing the auroral zones become narrower and move towards the poles as expected. With increasing *B* the discharge current increases, and hence the discharge voltage decreases due to the internal resistance of the rectifiers. In connection with this discharge another experiment was made, which may be interesting to note. A small non-inductive resistance (150 ohms) was coupled in series with the anode, and the noise voltage over this resistance was measured with an oscillograph, the amplifiers of which had a frequency band ranging from a low value up to 5 Mc/s. The result was, that no noise was observed for low magnetic fields up to about 425 gauss at the poles. At this field a noise suddenly appeared corresponding to a noise current with top value of about one or two μA. This noise was sensibly constant up to the highest magnetic fields. The anode current was 6.3 mA. It is probable that the noise level is much higher nearer the terrella, where space charge is accumulated, so that discharges along the magnetic field lines take place. The author hopes to be able to make more detailed investigations on this later.

In fig. 6 photos are shown of the 'invisible' discharge. As the electric field was much stronger than for the visible discharge, the magnetic field where the auroral zones begin to appear is much stronger, viz. 800 gauss.

According to MALMFORS (1946) the discharge starts at the singular point between the terrella and the cathode in the equatorial plane, where the drift velocity of the electrons is zero (the drift due to the crossed electric and magnetic fields opposes and equals that due to the inhomogeneity of the magnetic field). At this point the electrons spiral repeatedly around the same magnetic field line from the northern to the southern auroral zone thus causing an appreciable ionization. It should therefore be possible to switch off the discharge by a horizontal plate at this point. Such a plate made of insulating material was inserted, but it had no appreciable effect neither on the visible discharge nor on the invisible one. Then it was thought that the plate may be charged by

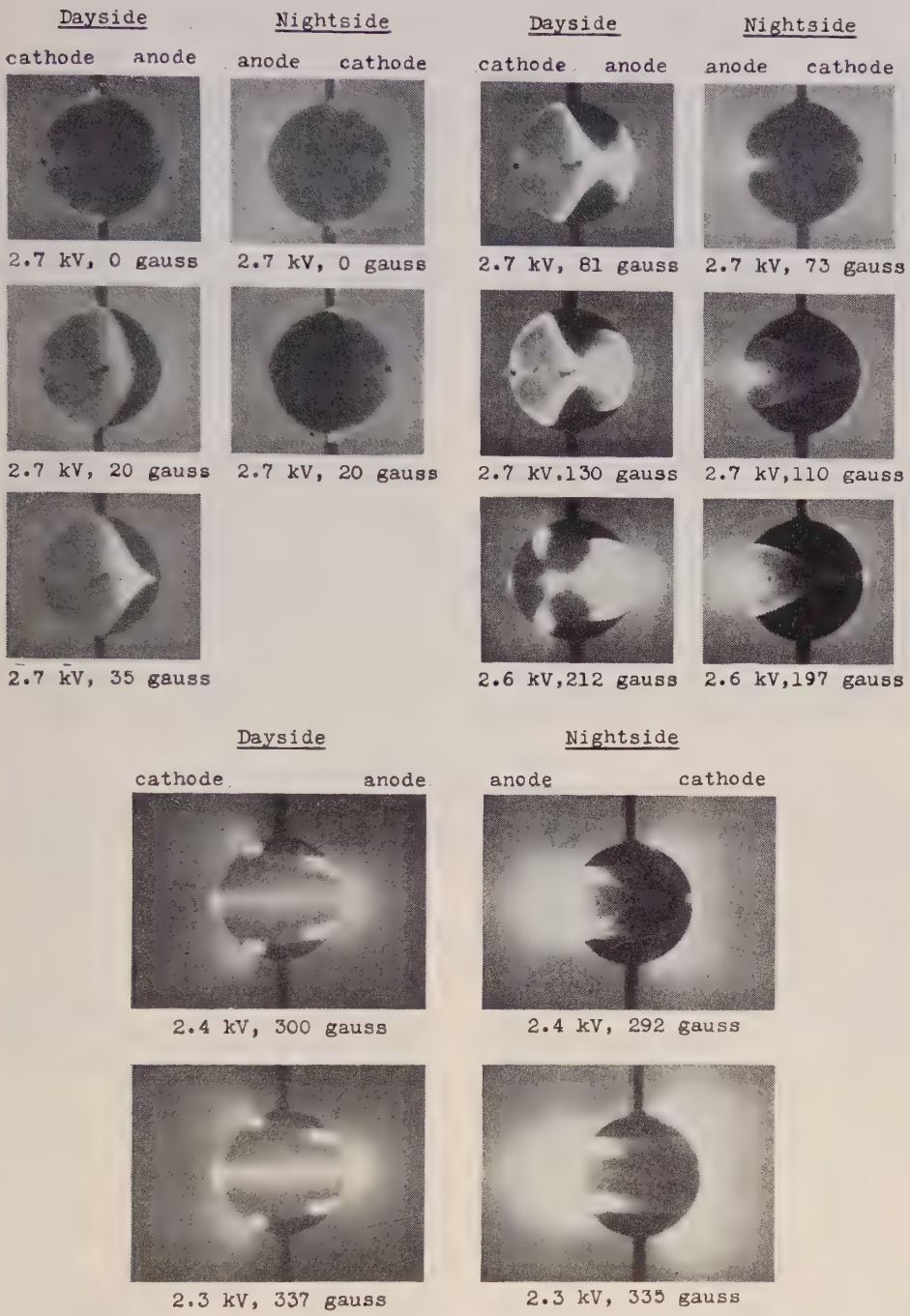


Fig. 5. Visible discharge. Below every picture the cathode voltage with respect to the terrella and the magnetic field at the poles of the surface of the terrella are given. Anode voltage was 30 volts. Pressure $3.5 \cdot 10^{-3}$ mm. Cathode current varying between 10 mA at 0 gauss and 22 mA at 1,350 gauss, anode current between 6.3 mA at 0 gauss and 5.5 mA at 1,350 gauss. The figure is continued on page 72 and 73.

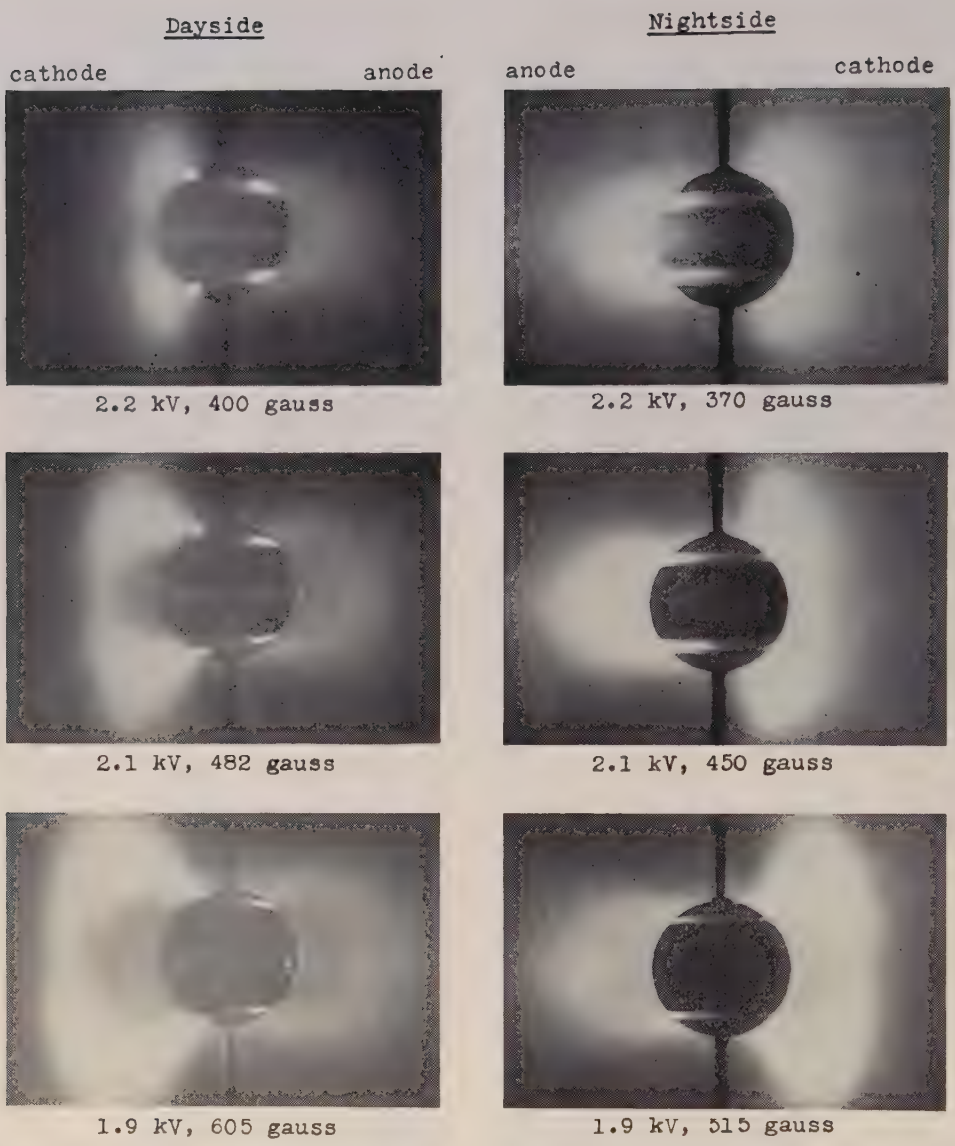


Fig 5. See text on page 71.

electrons to a negative potential thus reflecting the electrons on both sides, so that they could oscillate from the equator to the auroral zones. Therefore, the lower side of the plate was covered with a copper sheet, the potential of which could be varied. The effect of this is shown in figs. 7 and 8. The whole discharge is not switched off, but a part of the auroral zones is 'erased'. This is also sketched in fig. 12. The plate seems to have a greater effect in this respect on the visible than on the invisible discharge. One can, however, qualitatively understand that the singular point of zero drift velocity may have moved to another longitude because of the modification of the electric field caused by the plate. It seems as if the discharge

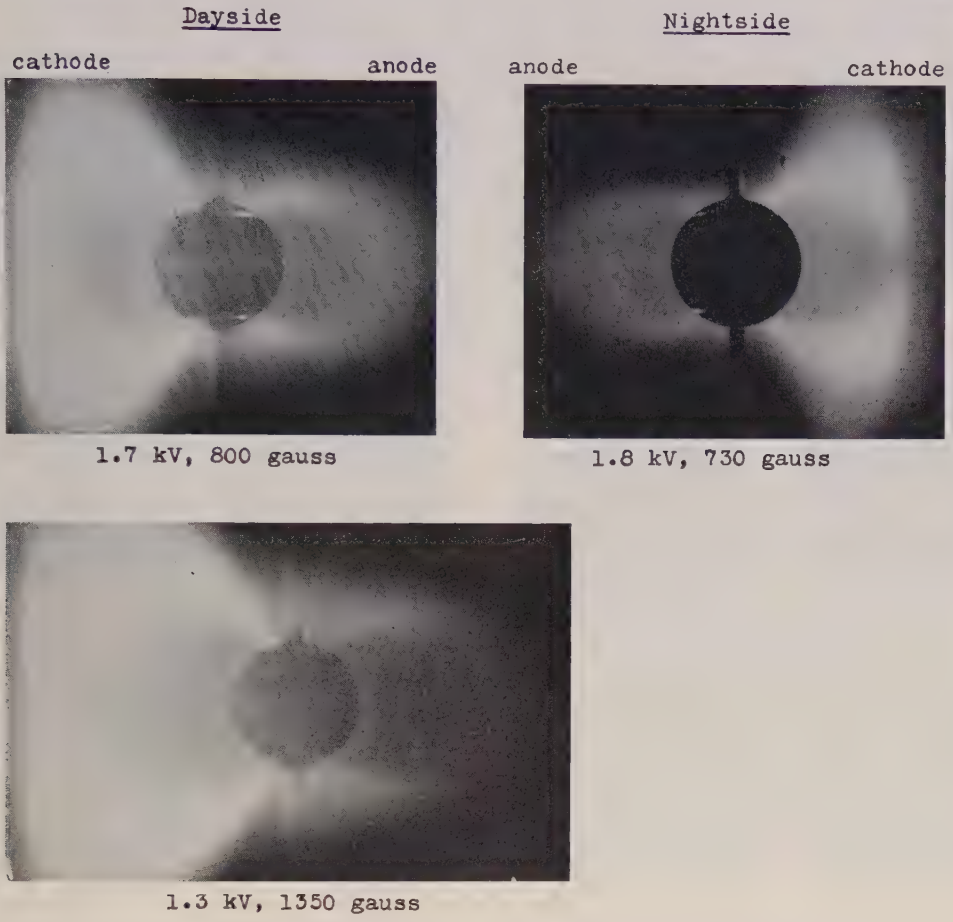


Fig 5. See text on page 71.

can always elude the investigator, but this experiment shows the great importance of the equatorial region for the generation of the auroral zones.

One may question if the two bright spots, one within each of the auroral zones on the cathode (morning) side of the terrella, have anything to do with the singular point of zero drift velocity, e.g. if they are the projection of this point on the terrella along the magnetic field lines. Fig. 9 shows that this is not the case, because the movable plate inserted between the cathode and the spot shadows the latter entirely. In the last photo of the figure a magnetic field line is indicated from the spot to the shadowing plate, showing that the electrons responsible for the spot move from the point, where a magnetic field line is perpendicular to the boundary between the plasma and the cathode fall. In the cathode fall the magnetic field may be too weak to deflect the electrons appreciably, but the plasma begins, where the electrons are caught by the field lines.

Fig. 10 shows ho the awuroral zones can be cut off by a plate in a meridional plane. Malmfors has also made this experiment, from which it is apparent that the auroral zones in some way are generated at the morning side

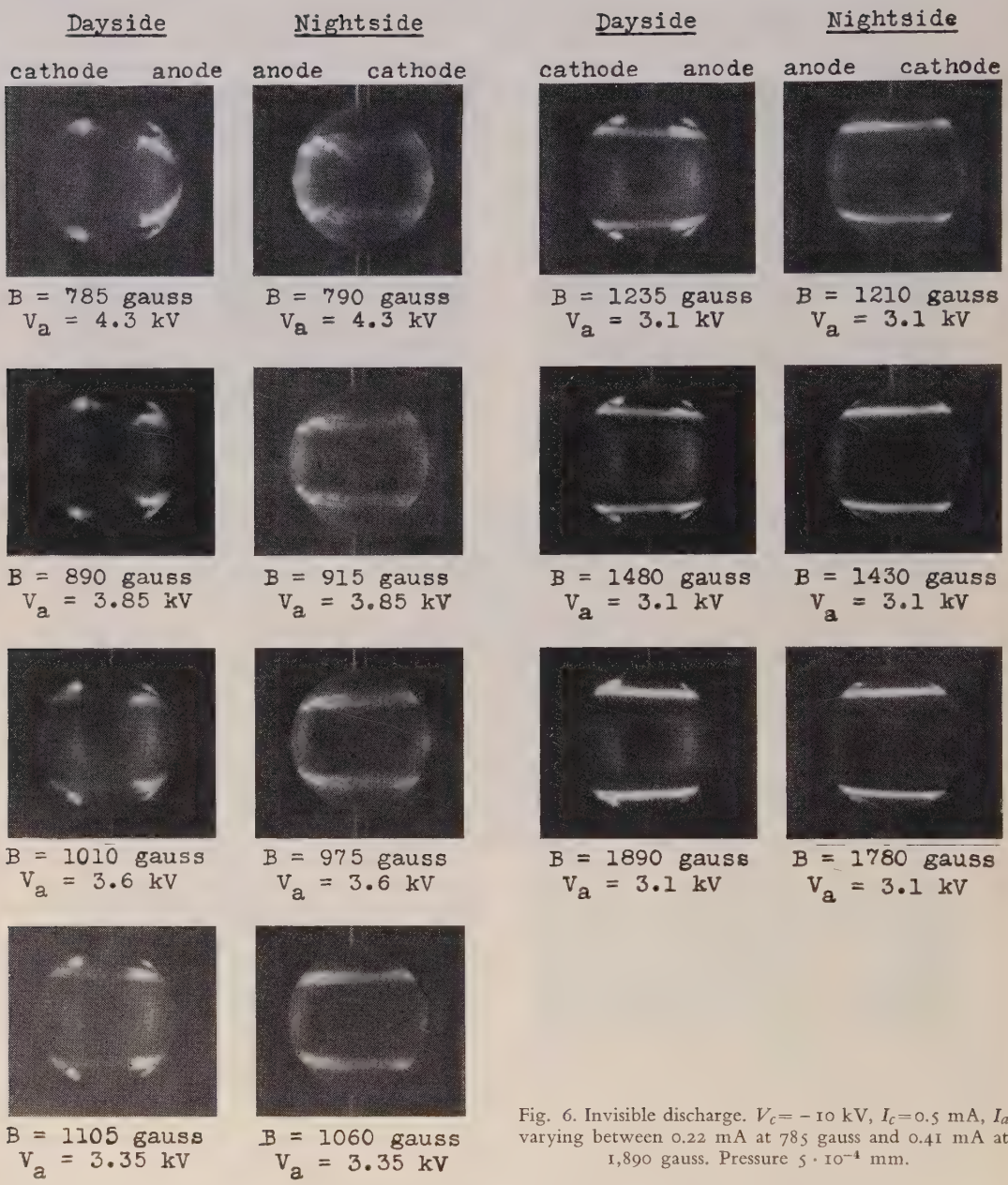


Fig. 6. Invisible discharge. $V_c = -10$ kV, $I_c = 0.5$ mA, I_a varying between 0.22 mA at 785 gauss and 0.41 mA at 1,890 gauss. Pressure $5 \cdot 10^{-4}$ mm.

of the terrella, even if the generation is not perhaps exactly of the same kind as Malmfors has proposed.

A plate of the form seen in fig. 11 was also used. The teeth on the lower edge of the plate made the auroral zone 'dashed'. It could easily be seen, that a tooth on the plate and the corresponding point on the auroral zone were situated in the same meridional plane as far as one could judge by eye.

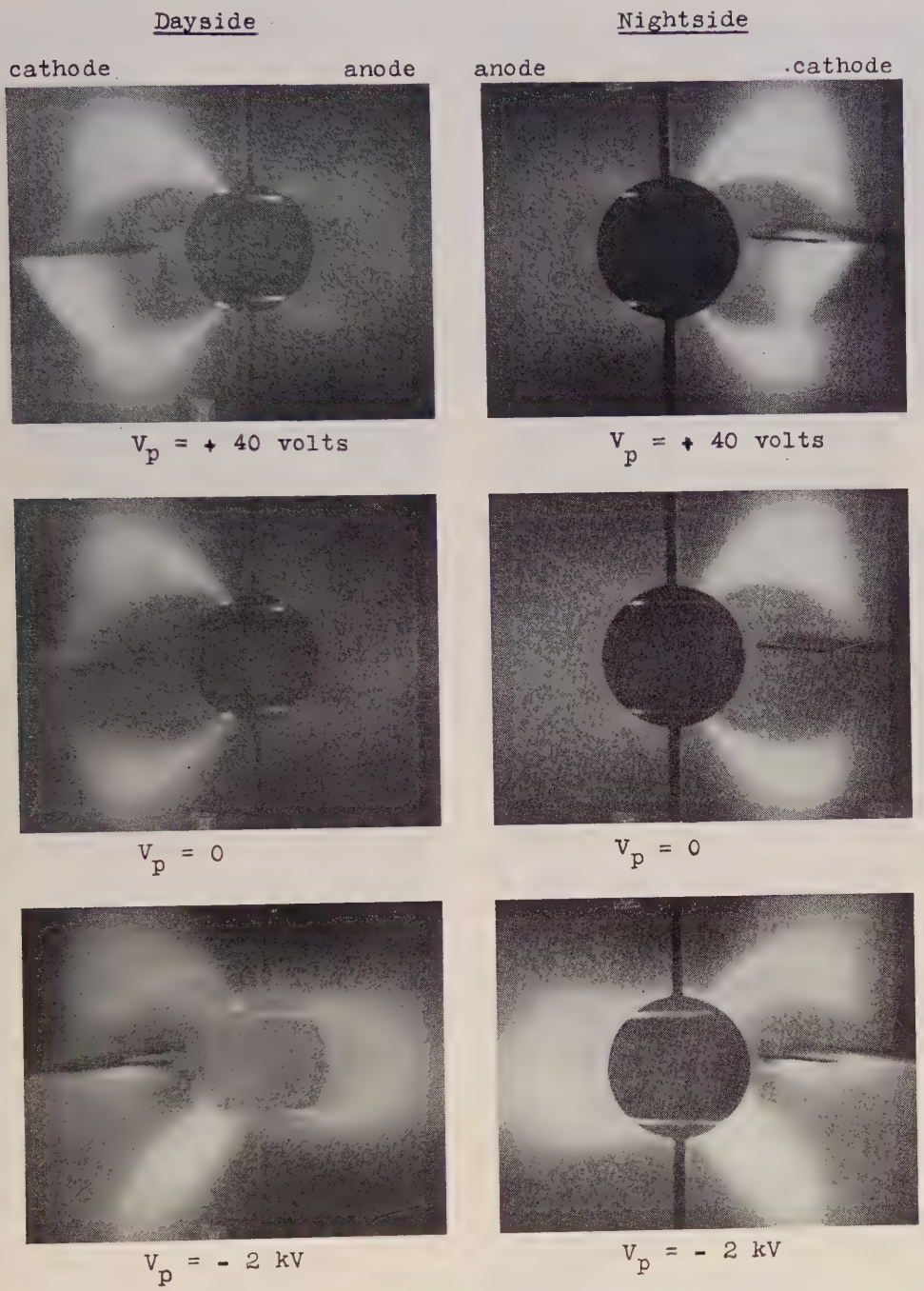


Fig. 7. Visible discharge. V_p = potential with respect to the terrella of the horizontal plate inserted on the cathode side. $V_c = -2$ kV, $V_a = +30$ V, $B = 1{,}090$ gauss, pressure $2.5 \cdot 10^{-3}$ mm.

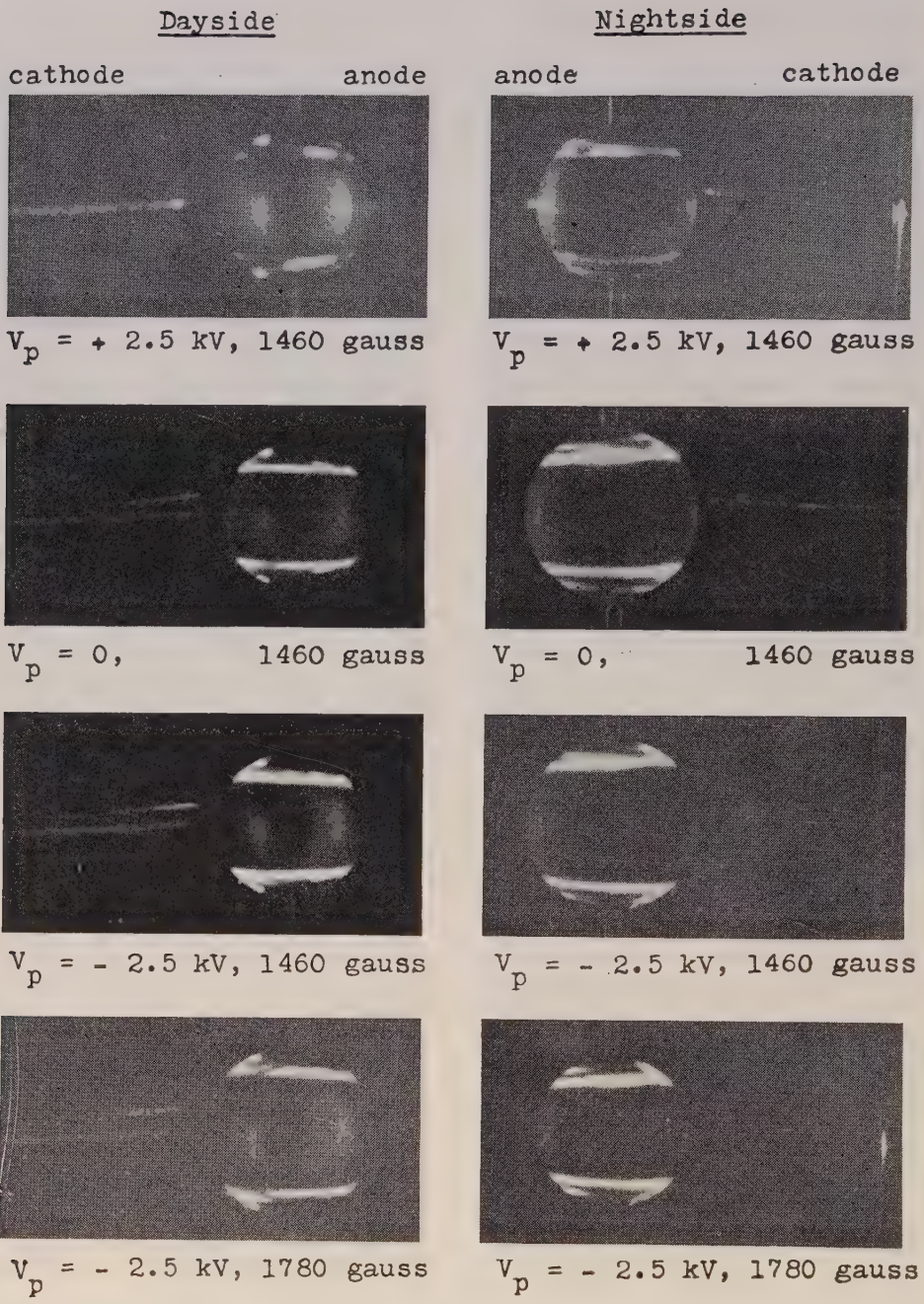

Dayside
Nightside
cathode
anode
anode
cathode
V_p = + 2.5 kV, 1460 gauss
V_p = + 2.5 kV, 1460 gauss
V_p = 0, 1460 gauss
V_p = 0, 1460 gauss
V_p = - 2.5 kV, 1460 gauss
V_p = - 2.5 kV, 1460 gauss
V_p = - 2.5 kV, 1780 gauss
V_p = - 2.5 kV, 1780 gauss

Fig. 8. Invisible discharge. $V_c = -10$ kV, $V_a = +5$ kV, pressure $5 \cdot 10^{-4}$ mm. Horizontal plate on the cathode side.

Calculation of forbidden zones according to Störmer and comparison with the experiments

One of the most important features of the discharges are the auroral zones and their latitude dependence on the magnetic and the electric field. STÖRMER (1933—37) has calculated the orbits of electrons moving from infinity into a magnetic dipole field and has found, that if the magnetic dipole field is strong enough or the electron energies small enough, there are forbidden zones resembling toroids around the dipole axis in the equatorial plane. It may be advantageous to calculate the size of these forbidden zones and to compare this with the forbidden latitudes on the terrella as seen on the photos. The reason for this is that there seem to be two possible explanations of the auroral rings on the terrella. The first is that electrons emitted from the cathode move in orbits very similar to those calculated by Störmer (properly modified by the electric field), thus causing the auroral zones. The second is that the particles impinging on the terrella are mainly formed by ionization in the discharge region and have then moved in orbits like those calculated by Alfvén. In the second case the energies of the impinging particles are lower than in the first case. We will therefore call the first kind of orbits high-energy orbits and the second kind low-energy orbits.

Störmer has defined a unit length C, the use of which makes all equations dimensionless.

$$C = \left(\frac{\mu_0}{4\pi} \cdot \frac{a}{B\varrho}\right)^{1/2} \qquad (16)$$

a = dipole moment,
$B\varrho$ is characteristic of the particle,
μ_0 = permeability of vacuum.

We confine ourselves to the non-relativistic case where $B\varrho$ is connected with the kinetic energy by

$$B\varrho = 3.37 \cdot 10^{-6} \cdot V^{1/2} \qquad (16)$$

V is expressed in volts, ϱ in metres and B in Vs/m². All lengths are measured in C-units so that the distance

$$r = AC \qquad (18)$$

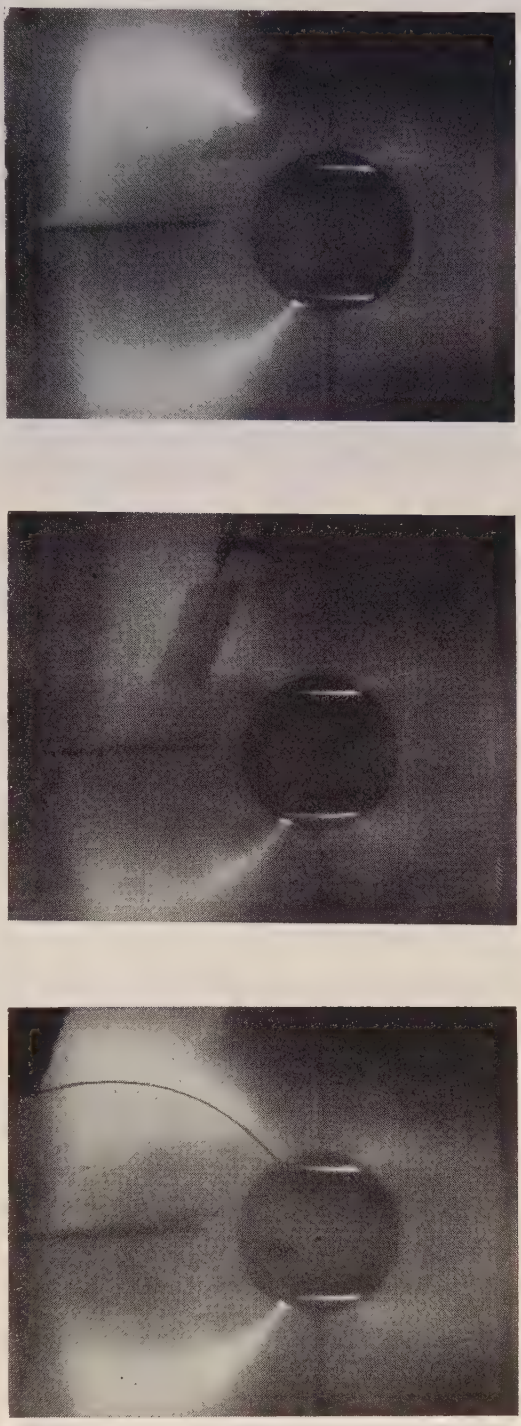
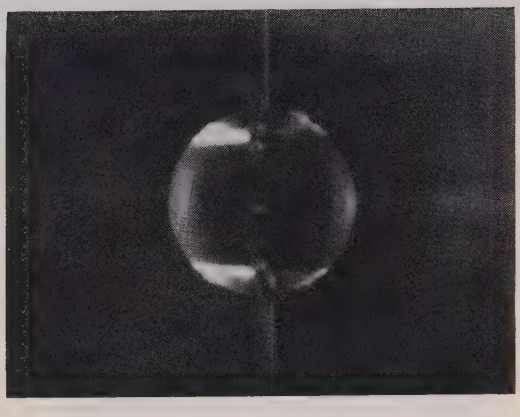
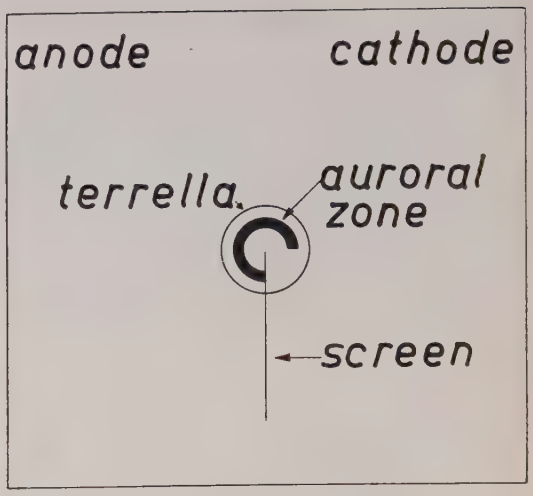


Fig. 9. Visible discharge. Dayside. Spot shadowed by movable plate. $V_c = -1.95$ kV, $V_a = +20$ V, $B = 1{,}110$ gauss, pressure $2.5 \cdot 10^{-3}$ mm.





a b

Fig. 10 a. Invisible discharge seen from the nightside. Auroral zone cut off by a vertical plate (screen) in a meridional plane. $V_c = -10$ kV, $V_a = +3.2$ kV, $B = 1{,}900$ gauss, pressure $5 \cdot 10^{-4}$ mm. b. Sketch of the experimental arrangement seen from above.

If the dipole moment a is given in terms of the magnetic field B_p at the poles of the terrella, we find, that if r is the distance from the dipole to the particle, A and V are connected by

$$V = 0.594 \cdot 10^8 \cdot B_p^2 \cdot A^4 \qquad (19)$$

We use Störmer's theorem

$$\cos \omega = \frac{2\gamma}{A \cos \varphi} + \frac{\cos \varphi}{A^2} \qquad (20)$$

φ = magnetic latitude,

Fig. 11. Visible discharge. Nightside. 'Dashed' auroral zone. $V_c = -1.95$ kV, $V_a = +20$ V, $B = 1{,}110$ gauss, pressure $2.5 \cdot 10^{-3}$ mm.

ω = the angle between the impact direction o the particle and the east-west direction,
γ is a constant depending on the angular momentum of the particle with respect to the dipole axis at infinity.

It is easy to find that the boundary of the forbidden zone, i.e. the lowest latitude where aurora can appear, is defined by $\omega = 0$, $\cos \omega = 1$. Thus we can solve for $\cos \varphi$ in (20).

$$\cos \varphi = \frac{1}{2} A^2 \left(1 \pm \sqrt{1 - \frac{8\gamma}{A^3}}\right) \qquad (21)$$

The particles can reach the dipole only if $-1 \leqq \gamma \leqq 0$ according to Störmer, and as $\cos \varphi > 0$, we can only use the positive sign in (21). We assume that there are electrons with every γ-value between 0 and -1 present in the discharge. We are only interested in the lowest possible latitude φ_0, that is the highest possible value of $\cos \varphi$, which occurs if $-\gamma$ is as large as possible. Thus we must put $\gamma = -1$ which gives

$$\cos \varphi_0 = \frac{1}{2} A^2 \left(1 + \sqrt{1 + \frac{8}{A^3}}\right) \qquad (22)$$

We see from this that the equator is forbidden only if the A-value corresponding to the radius of the terrella is less than $\sqrt{2} - 1 = 0.414$.

By using (19) and (22) we can calculate φ_0 as a function of V and B_p. If the electrons from

the cathode are accelerated in a relatively short cathode fall to an energy corresponding to the potential difference between cathode and terrella and then move only under the influence of the magnetic dipole field, then the lowest latitude of aurorae as seen from the photos should depend on V_c and B_p in the same way as φ_0 on V and B_p. This may in particular be the case in the visible discharge but not in the invisible one, where there is by no means a 'relatively' short cathode fall. It is also implied that no collisions and space charges appreciably affect the general behaviour of the discharge excepting their responsibility for the cathode fall.

Particularly in the invisible discharge the electrons do not immediately acquire an energy corresponding to V_c, but they are rather accelerated by the electric field through the whole path to the terrella. It is very difficult to calculate the forbidden zones if an arbitrary electric field is superposed on the magnetic field. STÖRMER (1931), however, has calculated the forbidden zones in the case of a spherically symmetric Coulomb field from the dipole. It is most advantageous to use his equation (l.c., p. 39, eq. 9)

$$c^2\left(1+\frac{c_1}{c^2}-\frac{b}{c^2}\cdot\frac{1}{r}\right)^2 \geq \left(\frac{c_2}{R}-a\frac{R}{r^3}\right)^2+c^2 \qquad (23)$$

r = distance from dipole to particle,
$R = r\cos\varphi$ = equatorial distance,
c = velocity of light,
c_1 = kinetic energy at infinity divided by the rest mass of the particle.
a and b are defined by

$$a=-\frac{e\mu_0 M}{4\pi m_0},\quad b=\frac{eV_0 r_0}{m_0} \qquad (24)$$

μ_0 = permeability of vacuum,
e = electronic charge in Coulombs,
m_0 = electronic rest mass in kg,
M = magnetic dipole moment in Am^2,
V_0 = potential in volts relative to infinity at r_0 metres from the dipole.
c_2 is a constant corresponding to γ in equation (20).

With a similar treatment of these equations as of (20) we get

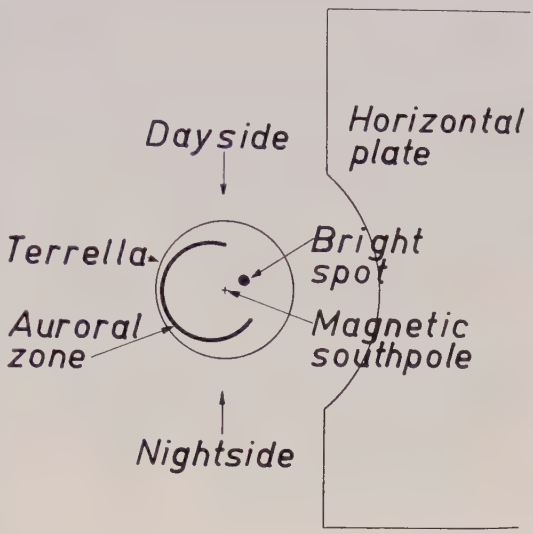


Fig. 12. Sketch of northern auroral zone seen from above. A horizontal plate is inserted between the terrella and the cathode so that a part of the auroral zone is 'erased'. Compare fig. 7.

$$\cos\varphi_0=\frac{1}{2}Z_0^{3/2}\left(1+\sqrt{1+\frac{3\sqrt[3]{16}}{Z_0^2}}\right) \qquad (25)$$

$$Z_0=r_0\left(-\frac{2b}{a^2}\right)^{1/3}=0.256\cdot 10^{-2}\left(\frac{V_0}{B_p^2}\right)^{1/3} \qquad (26)$$

where r_0 = terrella radius. It has been assumed that the kinetic energy at infinity ($c_1 m_0$) is zero. This result can be applied to our experiment if it is assumed, that V_0 is the potential difference between terrella and cathode and that the cathode is at infinity and the electrons start from the cathode with zero velocity. It may be of interest to compare this with the experimental results, although the electric field is not at all of the Coulomb type. The most important differences are, that in the experiments we have an anode, and the equipotential surfaces are not spheres, and thus they produce a radial component of $E\times B$, which means that the electrons can drift perpendicular to the magnetic field lines far closer to the dipole than if $E\times B$ had no radial component.

The results of these calculations are compared in figs. 13 and 14 with the experimental results obtained from the photos in figs. 5 and 6 respectively. It is to be expected that the experimental values should lie below the theoretical curve $\varphi_0(B_p)$ according to equations (25) and (26), where an electric Coulomb field

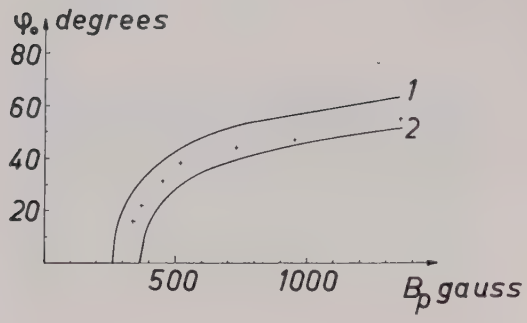


Fig. 13. The lowest allowed latitude φ_0 as a function of the magnetic field according to Störmer compared with the experimental values obtained from the visible discharge in fig. 5.

Curve 1: An electric Coulomb field is taken into account.
Curve 2: Neglecting the electric field.
Experimental values are indicated by +.

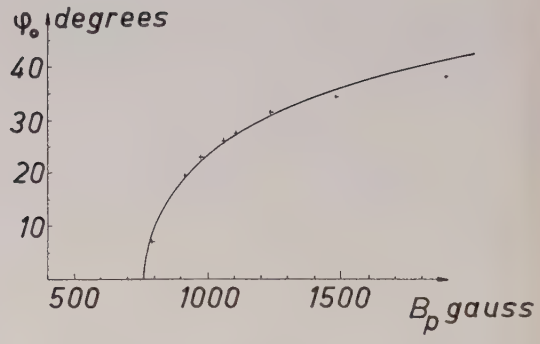


Fig. 14. The same as in fig. 13 but for the invisible discharge in fig. 6. No electric field is taken into account. Experimental values are indicated by +.

is also taken into consideration. This means that the electrons can be expected to reach lower latitudes than this theory permits because of the radial component of $E \times B$ in the experiment, and also because of low energy plasma electrons and ions moving according to Alfvén's theory. The latter will be considered in the next section.

It is seen from figs. 13 and 14 that our expectations are realized. In the case of the invisible discharge the experimental values for higher magnetic fields are even below the theoretical curve obtained from (22) without electric field. The curve calculated from (22) lies always below the one calculated from (25).

Alfvén's forbidden zones compared with the experiments

As already pointed out the auroral zones seen on the terrella may also be due to low energy plasma electrons moving according to Alfvén's calculations. In the case of an invisible discharge, however, one may not be allowed to speak of a real plasma, but still there are secondary electrons produced by ionization, perhaps mainly by the electron beam (see fig. 2), which may determine the latitudes of the auroral zones. In order to see if this is quantitatively possible, equation (7) or the last of equations (14) can be used. If we proceed in the same way as when we obtained (7 b), we find that the energy of the electrons responsible for the forbidden zone is

$$eV = e\frac{LE}{k^n} \tag{27}$$

at the distance kL from the dipole. The exponent n is 3 using (7) and 2 using (14). If (7) is used, V is the 'perpendicular' energy,

We first apply (27) to the visible discharge. assuming that the electrons and ions responsible for the auroral zones are formed by ionization mainly at a distance kL from the dipole. Since the terrella is situated in the plasma, the E-value prevailing there must be used. Assuming the anode voltage V_a relative to the terrella to be an approximative measure of the electric field in the plasma, we get

$V_a \approx 30$ volts,
$d \approx 30$ cm = distance from terrella to anode,
$E = V_a/d = 1$ V/cm.

L may be seen from the photos to be about 15 cm or 3 terrella radii and k can be of the order of 1 to 5. From (27) we then see that eV can be of the order of a few electron volts, which seems to be a quite reasonable energy for electrons formed by ionization in a plasma.

In the invisible discharge we cannot assume the starting energy of the secondary electrons to be only a few volts, because we now have so strong an electric field, that the energies are determined by equation (3), suitably modified for $V = V_\perp$ instead of $V'_\perp$

$$V = 2.85 \cdot 10^{-12} \left(\frac{E}{B}\right)^2 \tag{28}$$

A typical value of the discharge voltage $V_a - V_c$ is 14 kV, so that $E = 23$ kV/m, and if B is say 50 gauss $= 5 \cdot 10^{-3}$ Vs/m^2, we get $V = 60$

volts. Thus the best we can do is to combine (27) and (28) and solve for k. This can be done if we use the law of the dipole field

$$B = B_p \frac{R^3}{2r^3}, \qquad (29)$$

where B_p is the field at the poles of the terrella, and R is the radius of the terrella. Putting $r = kL$ we obtain

$$k^{6+n} = 8.8 \cdot 10^{10} \frac{B_p^2 \cdot R}{E\left(\frac{L}{R}\right)^5} \qquad (30)$$

If we take for example $B_p = 0.2$ Vs/m^2 = =2,000 gauss, we see from fig. 14 that φ_0 is about 38° experimentally, and the correspondnig magnetic field line cuts the equatorial plane at a radius of 1.6 terrella radii. According to Alfvén's theory this point corresponds to 0.74 L from the dipole if $n=3$ and 0.63 L if $n=2$. Therefore, we get

for $n=3$: $L/R = 1.6/0.74 = 2.2$, $k = 1.76$,
for $n=2$: $L/R = 1.6/0.63 = 2.5$, $k = 1.74$.
These values are quite reasonable.

It is not possible to obtain a relation between B_p and the latitude of the auroral zones from equation (30), since we do not know how k varies with L. However, MALMFORS (1946) has obtained such a relation based on Alfvén's theory. See also appendix II equation (7).

$$\cos \varphi_a = \left(285 \frac{ER^5}{a^2}\right)^{1/16}, \qquad (31)$$

φ_a = maximum latitude angle of the auroral zones,
E = electric field,
R = terrella radius,
a = magnetic dipole moment.
Putting $E = 23$ kV/m and $R = 5.3$ cm as before and using the relation between a and B_p we obtain

$$\cos \varphi_a = \frac{1.47}{\sqrt[8]{B_p}}, \qquad (32)$$

where B_p is measured in gauss.

In fig. 15 the experimental results obtained from the photos in fig. 6 are compared with equation (32). It should be pointed out, that this equation is only valid, if there are real auroral zones with the equatorial regions of the terrella in the forbidden space, because

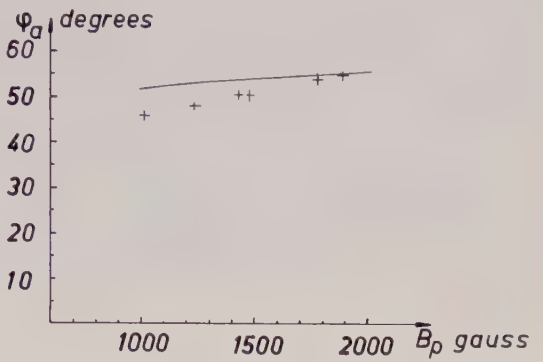


Fig. 15. The maximum latitude of the auroral zones according to Malmfors compared with the experimental values obtained from the invisible discharge in fig. 6. Experimental values are indicated by +.

the electric field theory is based on the assumption that the radius of curvature of the particles in the magnetic field is much smaller than the terrella radius. Therefore, the curves in fig. 15 are drawn only for $B_p > 1{,}000$ gauss. It is seen, that the experimentally obtained curve approaches satisfactorily the theoretical curve at high magnetic fields.

The angle φ_a is always greater than φ_0, the minimum latitude of the auroral zones, which we have considered in the preceding section.

The eccentricity of the auroral zones and Störmer's trajectories.

It may be seen from the photos, that the auroral zones are generally eccentric in qualitative agreement with Alfvén's theory. In this section we will examine this eccentricity in some detail. In the two preceding sections we have considered two possible explanations of the auroral latitudes and the forbidden zones. It is apparent, that from these considerations only, we cannot decide whether an Alfvén-mechanism or electrons moving in high-energy orbits are responsible for the auroral zones. With high-energy orbits we here, as before, mean trajectories described by high energy electrons coming directly from the cathode, and where space charge effects have been neglected. In the Alfvén-mechanism on the other hand, space charges are essential, because the auroral zones occur where space charge accumulated in the equatorial plane

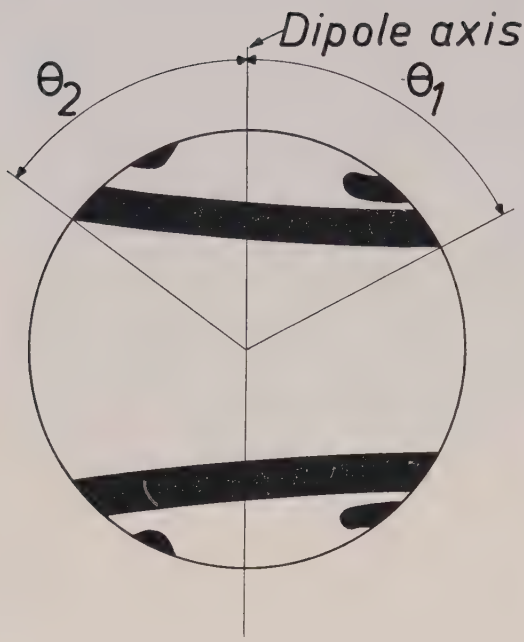


Fig. 16. Definition of the eccentricity of the auroral zones.

discharges along the magnetic field lines, and the energy of the particles may also be much smaller than in the case of high-energy orbits. As we shall see, the eccentricity strongly favours the Alfvén-mechanism, and there is also another reason against high-energy orbits.

We define the eccentricity ε as

$$\varepsilon = \frac{\Theta_1 - \Theta_2}{\Theta_1} \qquad (33)$$

The angles Θ_1 and Θ_2 are defined in fig. 16. It may be noticed that $\Theta_1 + \varphi_0 = 1/2\,\pi$. In this definition it has been assumed, that the border of the forbidden region in the equatorial plane according to the theory defines the lower edge of the auroral rings, but it may more probably define the middle of the rings. There are, however, often bright spots within the rings. We do not know whether they belong to the rings or not, and so it is difficult to find the middle of the rings. The lower edge is more well defined but seems to give something like 10 % lower values of ε than the middle of the rings. According to Alfvén's theory the eccentricity should be about 0.26 or 0.31 (because of the two possible ways of calculating the forbidden zones). The only value which relates to nature is obtained by interpreting the auroral curve from the measurements of the azimuth of auroral arcs. This gives $\varepsilon = 0.48$ (see Cosmical Electrodynamics p. 186 fig. 6.7), but this is a very uncertain value, and not much weight should be attached to it. From the invisible discharge we get a considerably smaller eccentricity. From fig. 17 it may be seen that at $B_p = 1{,}000$ gauss the eccentricity is about 0.13, and at increasing field it goes down to about 0.10. This discrepancy may be due to the geometry of the electric field. In the theory it is assumed to be homogeneous, but in the experiment it is strongly modified by the terrella. It has also been observed, that the eccentricity strongly depends on the terrella potential relative to anode and cathode, if the potential difference between anode and cathode is fixed. We can say that $\varepsilon = \varepsilon(V_a)$ with $V_a - V_c = V_d = \text{constant}$. The dependence is such that ε increases with V_a. The theory corresponds to $V_a = 1/2\ V_d$, but the curve in fig. 17 is drawn for $V_a = 0.29\ V_d$. The experiment therefore gives greater eccentricities than those in fig. 17. More recent experiments indicate that the eccentricity at $V_a = 1/2\ V_d$ is about 0.20. Details of these experiments will be given in a later communication.

In nature the electric field is modified by the earth's ionosphere, as it is in the experiment by the terrella. The above mentioned value of ε in nature is probably too high. This is indicated by magnetic storm measurements and will be considered in a later communication.

There is also another very important question about the eccentricity: Is it possible to obtain eccentric auroral zones from Störmer trajectories, and can they be continuous around the whole earth or terrella? A detailed study of Störmer's papers reveals that this is not the case. We can get a good picture of this by examination of his paper in Astroph. Norv. I. 4, second part of Pl. XI. The orbits are there shown in the R—z-plane, the dipole pointing in the z-direction. If we follow the different orbits, we see that for higher values of $\gamma_1 = -\gamma$, the electrons make an increasing number of oscillations through the equatorial plane and reach lower latitudes of the terrella. This is exactly in accordance with our previous calculations (see the discussion before equation (22)). The increasing number of oscillations means that the electrons move through an increasing meridional angle. This may be seen

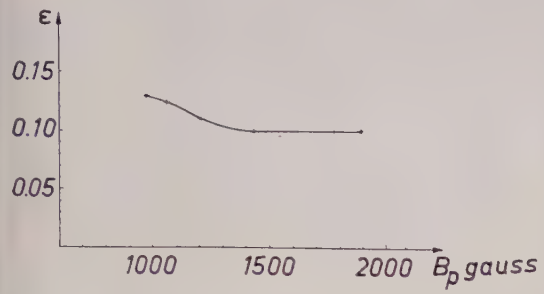


Fig. 17. The eccentricity of the auroral zones in the invisible discharge in fig. 6.

in Astroph. Norv. II. 1, Pl. III and the accompanying table on pp. 68—69 in the same paper. In fact, it is possible to construct from Störmer orbits an auroral zone, which is like a spiral starting from a high latitude on the cathode side and asymptotically attaining a latitude circle from above. This has been done by Störmer and is shown in fig. 18 with subsequent γ-values indicated. The observed auroral zone on the terrella may therefore be in agreement with Störmer's theory on the dayside, where its latitude is decreasing from left to right, but not on the nightside, where it is increasing from left to right. The spiral also shows, that it is not possible to obtain auroral zones which are continuous everywhere, especially not on the cathode side. In the experiment they are continuous at least in the case of invisible discharge, although we have also noted a bright spot within the rings on the cathode side, which probably is due to Störmer orbits (see fig. 9).

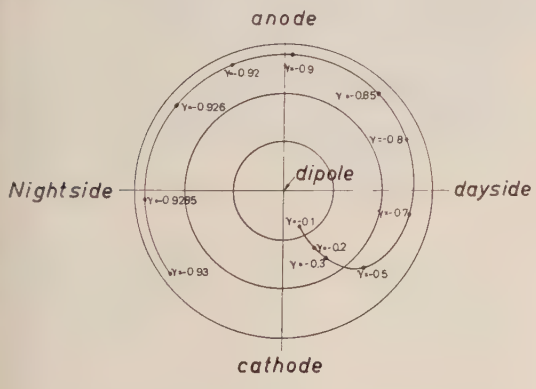


Fig. 18. 'Auroral spiral' constructed from Störmer orbits. See Störmer: Corpuscular theory of the Aurora Borealis, *Terr. Magn. and Atmosph. El*, **22**, fig. 4, p. 32.

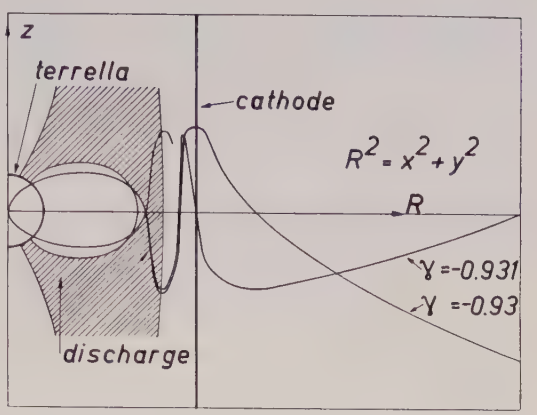


Fig. 19. Comparison of Störmer orbits in the R—z-plane and the visible discharge in the last photo in fig. 5. The Störmer orbits are taken from his paper in Astroph. Norv. I, 4, Pl. XI.

In fig. 19 some Störmer orbits from Astroph. Norv. I. 4, Pl. XI are compared with the discharge in fig. 5, the last dayside photo. It is very clearly seen, that there is insufficient space in the discharge for the electrons to oscillate through the equatorial plane so many times, that they can reach the nightside of the terrella (compare the γ-values in fig. 19 with those in fig. 18).

From this we may conclude, that the general behaviour of the discharge is not governed by high-energy orbits, although some subsidiary features, such as the bright spots on the cathode side, probably are due to such orbits.

It is apparent, that Alfvén's conception has been confirmed by these experiments, that space charge accumulated in the equatorial plane, rather than destroy the whole pattern of motion, immediately discharges along the

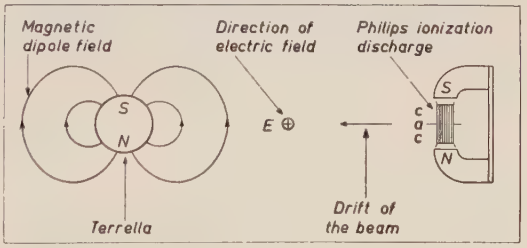


Fig. 20. Arrangement for suggested experiments where the neutral beam from the sun is also simulated. 'N' and 'S' refer to magnetic north and south poles. At the 'Philips' discharge, 'c' and 'a' mean cathode and anode respectively.

field lines. One should bear in mind, that experiments like these can never be a proper scale model of nature, but they can show in a convincing way, if a certain proposed mechanism, such as Alfvén's 'auroral mechanism', is likely to exist or not.

Suggestions for further experiments

As already pointed out it is impossible to scale down the real auroral discharge to laboratory dimensions in an entirely satisfactory way. In this section we will discuss one possibility of improving the experimental arrangement, so that we can be more sure of closer similarity to the real conditions in nature, as suggested by the electric field theory.

The neutral beam from the sun constitutes the most important difference between nature and experiment. Therefore, experiments arranged as indicated in fig. 20 may be more like nature. In 'Philips-type' discharges electrons and ions are produced, and they drift towards the terrella because of the crossed electric and magnetic fields. As will be seen, this arrangement makes it possible to work with lower electric field and without the usual type of discharge between the plates. The only purpose of the electric field is then to guide the drift of the particles, so that the beam itself can produce space charges, which may cause discharges along the magnetic field lines, just as in the theory.

We have pointed out in the beginning of this paper, that according to the correct transformation laws we should have
$V_d \approx 100$ kV,
$B_p \approx 60$ million gauss $= 6{,}000$ Vs/m^2.
Using (11), however, we can get the same electron orbits (assuming the kinetic energies to be proportional to the applied voltage), if we use another magnetic field B' and another voltage V'_d, which are connected by

$$\frac{V'_d}{10^5} = \left(\frac{B'}{6{,}000}\right)^2 \tag{34}$$

If $B' = 0.2$ Vs/m$^2 = 2{,}000$ gauss, we get $V'_d = 10^{-4}$ volts. As 10^{-4} volts corresponds to an electron temperature of 1° K, this is impossible to obtain in any experiment. The only way to get any similarity with nature is to confine ourselves to a correct value of L according to (14) or (15)

$$L^3 = \frac{1}{E}\left(\frac{\mu_0}{4\pi}\varkappa\right)^{2/3}, \tag{35}$$

$$L^4 = \frac{\mu_0}{4\pi}\frac{a\mu}{eE} \tag{36}$$

If we want $L = 4r_0$ (r_0 = terrella radius) at $B_p = 2{,}000$ gauss, and if we use

$$\frac{\mu_0}{4\pi} a = \frac{1}{2} B_p \cdot r_0^3 \tag{37}$$

$$E = \frac{V_d}{5.4\, r_0}, \tag{38}$$

we get

$$\varkappa^{2/3} = 56\, V_d \tag{39}$$

or

$$\frac{\mu}{e} = 480\, V_d \tag{40}$$

respectively. If the magnetic field in the 'Philips' discharges is say 300 gauss the electronic energies in them must be

$$eV = e \cdot 56 \cdot V_d (3 \cdot 10^{-2})^{2/3} \approx 6\, eV_d, \tag{41}$$

or

$$eV_\perp = e \cdot 480\, V_d \cdot 3 \cdot 10^{-2} \approx 14\, eV_d. \tag{42}$$

In the 'Philips' discharges $V_\perp$ can be at most some hundred volts, and so V_d becomes less than 100 volts. This low value of V_d cannot cause a usual type of discharge between the plates.

Appendix I. Calculation of forbidden zone assuming the electronic velocities to be isotropically distributed everywhere.

According to the perturbation method for calculation of electron orbits (see Alfvén: Cosmical electrodynamics, Ch. II) we have

$$\mu = \frac{eV_\perp}{B} = \text{constant of the motion} \tag{1}$$

In a discharge collisions and plasma oscillations frequently occur, and they may to a certain extent maintain an isotropic distribution of the kinetic energies, which would not be possible if (1) is strictly valid. We may therefore assume, that an electron first moves a short distance, where (1) is valid. The magnetic field changes with a small value dB during

this motion, and according to (1) $V_\perp$ changes to $V_\perp + dV_\perp$, so that

$$\frac{V_\perp}{B} = \frac{V_\perp + dV_\perp}{B + dB} \tag{2}$$

Then collisions and plasma oscillations redistribute this energy change so that the isotropic distribution is maintained.

$$\frac{\frac{2}{3}V}{B} = \frac{\frac{2}{3}V + Vd}{B + dB} \tag{3}$$

where $2/3\ V = V_\perp$ but $dV = dV_\perp$ of equation (2), since $dV_\parallel = 0$ there. Integrating (3) we get the new constant of the motion

$$\varkappa = \frac{V^{3/2}}{B} \tag{4}$$

Using the same notation as earlier in this paper, e. g. in fig. 3, we have

$$V = V_0 + E(x_D - x), \tag{5}$$

$$B = B_0 + \frac{b}{(x^2 + y^2)^{3/2}} \tag{6}$$

with

$$b = \frac{\mu_0}{4\pi} a \tag{6a}$$

B_0 is the solar magnetic field and eV_0 is the electronic energy far away from the earth. Together with (4) this gives

$$\varkappa\left(B_0 + \frac{b}{R^3}\right) = \left(V_0 \quad E(x_D - x)\right)^{3/2} \tag{7}$$

Since B_0 and V_0 are very small compared to the other quantities, we can neglect them here. They affect only the orbits far away from the dipole, where they are of little interest. Then we get

$$\frac{\varkappa b}{R^3} = \left(E(x_D - x)\right)^{3/2}, \tag{7 a}$$

which is equivalent to

$$x_D - x = \frac{L^3}{R^2}, \tag{8}$$

$$L^3 = \frac{1}{E}(b\varkappa)^{2/3} \tag{9}$$

From this we get one of the curves in fig. 4 which cuts the x-axis at

$$x'_d = \sqrt[3]{2}\, L = 1.26\, L,$$

$$x'_m = -\frac{1}{2}\sqrt[3]{2}\, L = -0.63\, L.$$

At infinity we get (still neglecting V_0 and B_0)

$$x = x_D = \frac{3}{2}\sqrt[3]{2}\, L = 1.89\, L.$$

The y-axis is cut at

$$y = \pm\frac{\sqrt[3]{2}}{\sqrt{3}} = \pm 0.73\, L.$$

Finally we must estimate the error made by neglecting V_0 and B_0. By putting

$$\varkappa B_0 = V_0^{3/2} \tag{10}$$

into (7) we get

$$x_D - x = \frac{1}{R^2}\left(\frac{L}{\phi}\right)^3 \tag{11}$$

with

$$\phi^3 = (1+\delta)\left[1 - \left(\frac{\delta}{1+\delta}\right)^{3/2}\right]^{2/3}, \tag{12}$$

$$\delta = \frac{V_0}{E(x_D - x)} \tag{13}$$

Supposing $B_0 = 0.7 \cdot 10^{-5}$ gauss, $E = 1.4 \cdot 10^{-3}$ V/m and $L = 7r_0$, where r_0 = earth's radius, we get

$$\frac{V_0}{E} = 3.85 \cdot 10^5\, m = 0.86 \cdot 10^{-2}\, L. \tag{14}$$

If we use this value in (13) and put $x_D = 1.89\ L$, we find that $\phi = 1.026$ at $x = 1.8\ L$. Since $\phi > 1$ and steadily increasing with increasing x for every $x < x_D$, the error made by neglecting V_0 and B_0 is negligible.

Appendix II. Calculation of the auroral latitudes according to Malmfors.

Earlier in this paper frequent reference has been made to Malmfors' experiments. Equation (31) is taken from his paper. Since it is out of print and somewhat inaccessible, it seems desirable to derive that equation here.

Electrons are assumed to oscillate along magnetic field lines from the northern to the southern hemisphere of the earth. If the

amplitude is large enough, the electrons hit the atmosphere and are absorbed there, thus causing aurora. If r_0 is the distance from the dipole to the equatorial point cut by the field line, we get the equation of this field line

$$r = r_0 \cos^2\varphi, \tag{1}$$

and the magnetic field strength is

$$B = \frac{\mu_0 a}{4\pi r^3}\sqrt{1+3\sin^2\varphi} = \frac{\mu_0 a}{4\pi r^3}\cdot m \tag{2}$$

Only the morning side of the earth (region between terrella and cathode) is considered, because the discharge is assumed to be generated there (at the singular point of zero drift velocity). The electric field E V/m is then situated in the same meridional plane as the magnetic field line. If $r=r_1$, $\varphi=\varphi_1$ is the turning point of the electron, the 'perpendicular' energy there is

$$eV_1 = eV_0 + eE(r_0 - r_1\cos\varphi_1), \tag{3}$$

where eV_0 is the 'perpendicular' energy in the equatorial plane. Since φ_1 is rather near 90°, $r_1 \ll r_0$ and V_0 is small, we can simplify (3) to

$$V_1 \approx Er_0 \tag{3a}$$

We also have

$$\frac{V_1}{B_1} = \frac{V_0}{B_0}. \tag{4}$$

According to (28) in the beginning of this paper V_0 is determined by

$$V_0 = 2.85\cdot 10^{-12}\frac{E^2}{B^2}. \tag{5}$$

Malmfors has used 5.7 instead of 2.85, becaues he has not considered the difference between $V_\perp$ and $V'_\perp$ discussed after equation (8) earlier in this paper. Combining (2), (3 a), (4) and (5) we arrive at

$$m^2\cdot r_1^3 = 285\frac{E}{a^2}r_0^8. \tag{6}$$

If the auroral zones are due to electrons with $r_1 \leqq R =$ earth's radius, and if φ_a is the maximum latitude angle of the auroral zones, we get with (1) and (6)

$$m^2\cdot\cos^{16}\varphi_a = 285\frac{ER^5}{a^2} \tag{7}$$

Since m is of the order of unity, this equation is equivalent to the earlier equation (31).

Acknowledgements.

The author is greatly indebted to Professor H. Alfvén, who has suggested these experiments and guided the work and stimulated it by means of helpful and most inspiring discussions. Dr. N. Herlofson has given invaluable advice and has made most interesting comments during numerous discussions. Mr A. Burger has constructed some parts of the apparatus and has assisted in bringing it into work. Mr H. Witt has assisted in performing many of the experiments and has also carried out most of the photographic work. Mr Spångberg has drawn the figures. Dr. T. S. Robinson has corrected the English. The author wishes to thank all these for their kind help. The work has been made possible by financial assistance from Statens Naturvetenskapliga Forskningsråd.

REFERENCES

Alfvén, H., 1939: A Theory of Magnetic Storms and Aurorae, I, *Kungl. Sv. Vetenskapsakad. Handl.* (3), **18**, No. 3.

— 1940: A Theory of Magnetic Storms and Aurorae, II, III, ibid. **18**, No. 9.

— 1950: Cosmical Electrodynamics, *Oxford.*

— 1955: On the Electric Field Theory of Magnetic Storms and Aurorae, *Tellus*, **7**, No. 1.

Cobine, J. D., 1941: Gaseous Conductors, *New York.*

von Engel, A. and Steenbeck, M., 1932: Elektrische Gasentladungen, *Berlin.*

Malmfors, K. G., 1946: Experiments on the Aurorae, *Ark. f. mat., astr. o. fysik*, **34B**, No. 1.

Meinel, A. B., 1951: Doppler-shifted Auroral Hydrogen Emission, *Astroph. Journal*, **113**, No. 1. See also *Rep. on Progr. in Phys.* **XIV**, pp. 121—146.

Störmer, C., 1931: Ein Fundamentalproblem der Bewegung einer elektrisch geladenen Korpuskel im kosmischen Raume, *Zeitschrift für Astrophysik*, **3**, No. 1.

— 1933—37: On the Trajectories of Electric Particles in the Field of a Magnetic Dipole with Applications to the Theory of Cosmic Radiation, *Avh. Vid. Akad. Mat.-nat. Kl., Oslo*, 1933—34; *Astroph. Norv., Oslo*, **I**, No. 1 and 4, 1934—35, **II**, No. 1 and 4, 1936—37.

Upper Air Temperatures over an Antarctic Station

A preliminary note by NILS JØRGEN SCHUMACHER

Norsk Polarinstitutt, Oslo, Norway

(Manuscript received July 27, 1954)

Abstract

Typical features of antarctic upper air temperature distributions are demonstrated by means of average values and selected ascents. A great seasonal change in the height of the tropopause is pointed out. Attention is drawn to the remarkable agreement between observed stratospheric temperatures in winter and Dobson's estimated values. The causes for the heating of the stratosphere in late winter and spring are discussed. The difference between the tropospheric temperatures in summer between Arctic and Antarctic is assumed to be due to the different value of the surface albedo. Rough estimates indicate that the absorption in the 9.7 μ band by ozone is of great significance to the observed stratospheric temperature contrast between the two polar regions in winter.

The aerological observations at Maudheim (lat. 71° 03′ S., long. 10° 56′ W.), the base of the Norwegian-British-Swedish Antarctic Expedition, 1949—52, started on March 15, 1950, well over a month after the landing on February 10. During March, quite a few soundings were carried out, but the quality of these first soundings was rather unsatisfactory, due to instrumental and operational difficulties. As from April 1, however, the programme of one sounding a day was steadily adhered to, whenever weather permitted launching. The daily ascents were suspended for 10 days in May, 1950, when weather conditions enforced a radical change in the launching technique hitherto used. The last sounding was made on January 10, 1952.

During the entire period some 650 sounding balloons were launched, nearly all carrying a radiosonde of Finnish design (Väisälä type) and a rawin transmitter for upper wind measurements.

In this note we shall deal with the temperature measurements only, which are now partly worked up. The upper wind data are being examined by Gordon de Q. Robin, phycisist of the expedition, who was responsible for this part of the investigations.

Table I contains the numbers of temperature data at stated pressures for the different months, both years combined. Thus, the table shows the number of observations upon which our mean values are based. All unreliable ascents have been disregarded, and there are only very few cases with more than one sounding a day.

All ascents have been recalculated after our return, and for higher levels corrections for radiation error have been applied according to the latest method introduced by VÄISÄLÄ

Table 1. Number of temperature readings at different standard pressure surfaces

	Jan.	Febr.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
850	38	23	31	51	46	58	60	60	56	58	57	58
700	37	23	31	51	44	58	59	60	56	58	57	58
600	37	23	31	51	43	58	59	60	56	58	57	57
500	37	23	30	51	42	58	58	60	56	57	56	57
400	37	23	30	51	42	57	55	60	55	57	54	57
300	37	23	30	51	42	57	53	59	54	57	52	57
250	36	23	29	51	41	56	53	59	54	57	52	56
200	36	23	29	46	38	56	52	59	54	57	51	56
150	35	23	28	42	36	54	46	57	53	56	50	53
100	35	23	25	38	32	45	27	46	49	54	45	48
85	27	21	19	30	25	30	9	23	40	49	38	42
70	22	19	18	22	21	9	3	18	35	44	34	39
60	20	17	17	18	11	4	1	6	26	38	25	35
50	16	14	12	12	10	2	0	2	21	29	17	23
40	9	9	8	7	2	0	0	0	12	19	10	14
30	3	3	2	3	0	0	0	0	6	13	2	2
25	2	0	0	1	0	0	0	0	4	7	1	1
20	1	0	0	0	0	0	0	0	1	1	0	0

(1941) and RAUNIO (1951). In this connection it should be mentioned that in summer the ascents were made in the evening with low solar altitude. In this manner we tried to reduce the radiation error to a minimum.

Some features of the upper air temperature distribution, according to selected ascents

Figure 1 shows six selected ascents from Maudheim, taken at different seasons. An ordinary log p, T diagram has been used. Ascent I shows typical summer conditions. At this time of the year there exists no ground inversion during the day, though at night a rather small inversion may develop. Even in the lowest layers the temperature is generally below 0°C, a value barely exceeded in very few cases.

A marked tropopause is usually found in summer, often rather close to the 300 mb level. Above the tropopause there is found a layer of variable thickness with a pronounced negative lapse-rate. Higher up the temperature increases slightly up to the highest levels reached.

Even though the day to day variation of the temperature has not been fully investigated so far, it is evident that these variations are comparatively small during the summer, especially in the stratosphere. If we consider the 35 readings obtained from the 100 mb level during January the extreme temperatures recorded are −38,2° C and −45,0°C, and in

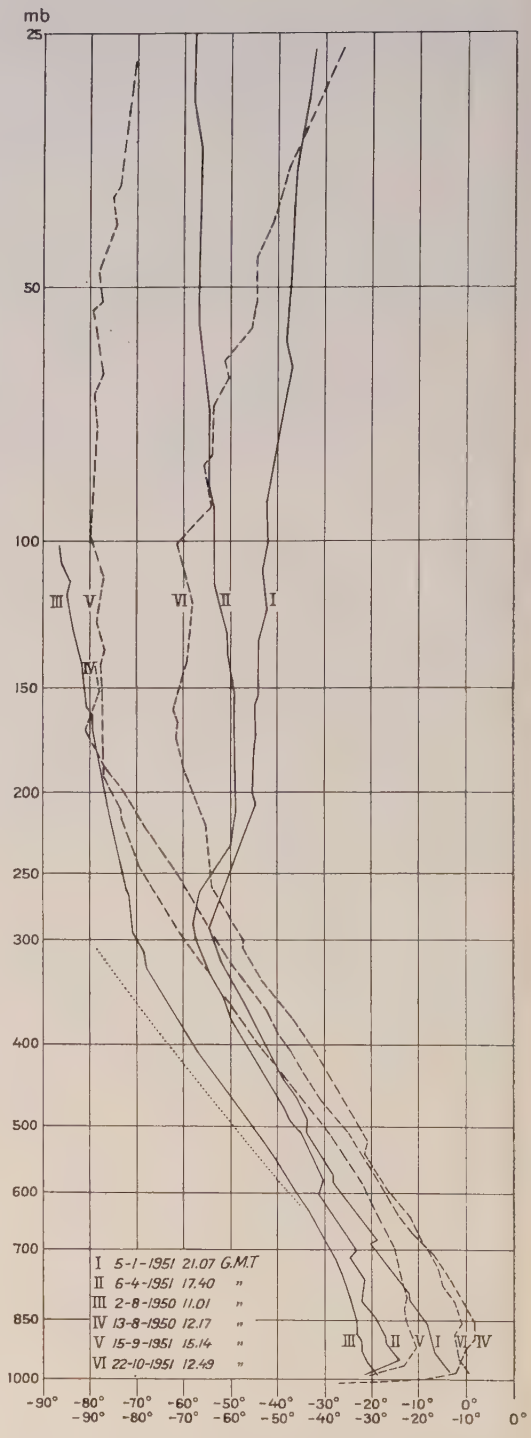


Fig. 1. Selected ascents. Full lines upper temperature scale, dashed lines lower temperature scale. Dotted line dry adiabatic.

21 of the cases the readings range between −41,0°C and −42,9°C.

In autumn, a ground inversion is frequently present even in the middle of the day, and the tropopause is usually very marked. The stratosphere has now become considerably colder, especially at the higher levels, and the temperature decreases slightly upwards (Ascent II).

Ascents III and IV show winter conditions: Generally a pronounced ground inversion is found, but a long-lasting gale or a heavy overcast sky may cause the inversion to disappear completely even in the middle of winter.

At this time of the year, temperatures in the lowest part of the troposphere undergo large day-to-day fluctuations, partly caused by vertical motions. At the top of the inversion layer or at a height of about 1,500 m, subsidence above may cause the temperature rise nearly to the highest values that are found in summer time. Ascent IV showed a temperature of −8,2°C at the 850 mb level, whereas the absolutely highest value measured at this level was −4,9°C on January 20, 1951.

One of the most interesting features of the antarctic winter ascents is the indefinite character of the tropopause, frequently observed during the winter months. COURT (1942) who first discussed this fact, goes as far as to say that the tropopause disappears during the antarctic winter. However, this is a matter of definition. In cases where no abrupt change in lapse-rate takes place, the generally adopted definition is: "The tropopause is taken as the point where the mean lapse-rate to the kilometre next above is 2°C or less, provided that it does not exceed 2°C for any subsequent kilometre", (Gibbs, Gotley and Martin, 1950).

As in the coldest months the lapse-rate usually exceeds 2°C at the highest levels of the soundings, the tropopause certainly disappears completely according to this definition. On the other hand, it is usually possible to point out a level above which the lapse-rate is quite different from the values found below, especially when mean values are considered. Therefore, to suit antarctic conditions, the quoted definition should be modified, and this can be done by deleting the subsidiary clause, the definition thus being: "The tropopause is taken as the lowest point in the transitional zone, where the mean lapse-rate for the kilometre next above is 2°C or less." This definition has been adopted for the Maudheim ascents, even when a real inversion exists. In this way it has been possible to define a tropopause in nearly all cases, often, however, with poor distinctness (Ascent III). Still, in some few cases the lapse-rate exceeds the limit mentioned for all layers between the 700 and the 100 mb surface, the tropopause thus completely disappearing even according to our definition.

On the other hand, the tropopause may occasionally be very distinct, even in the middle of the winter. In such cases it is usually found at considerable heights, and with extremely low temperatures, possibly the lowest values ever recorded in the atmosphere, whereas the lower part of the troposphere is warm and dry. In other words, these ascents show conditions similar to those observed in warm anticyclones on middle latitudes, the absolute temperature values being disregarded.

Although the extremely low temperatures in the stratosphere usually are connected with high temperatures in the lower troposphere, this is not always the case, as seen from ascent V. Furthermore, this ascent shows that very low temperatures may occur as late as in the middle of September. In late spring the stratosphere gradually becomes warmer, and as the heating starts at the highest levels, the lapse-rate may obtain comparatively great negative values. Thus, the temperature may rise as much as 30 to 40 degrees centigrade from the 100 mb to the 25 mb level, Ascent VI.

Mean temperature distribution

Mean ascent curves have been derived for half monthly as well as for monthly periods. In the lowest layer, the mean temperatures are determined for heights of 250, 500, 750 and 1,000 metres. Higher up the means refer to isobaric surfaces. Starting at 850 mb, the distances used between successive reference surfaces were 30, 20 and 10 mb for the layers 850 to 640 mb, 640 to 300 mb and 300 to 100 mb, respectively. Above the last mentioned surface a constant distance of 5 mb has been used. In this way, a smooth mean ascent curve has been obtained for each period. Furthermore, it has been possible to make use of additional readings above the highest standard surface reached by the balloon.

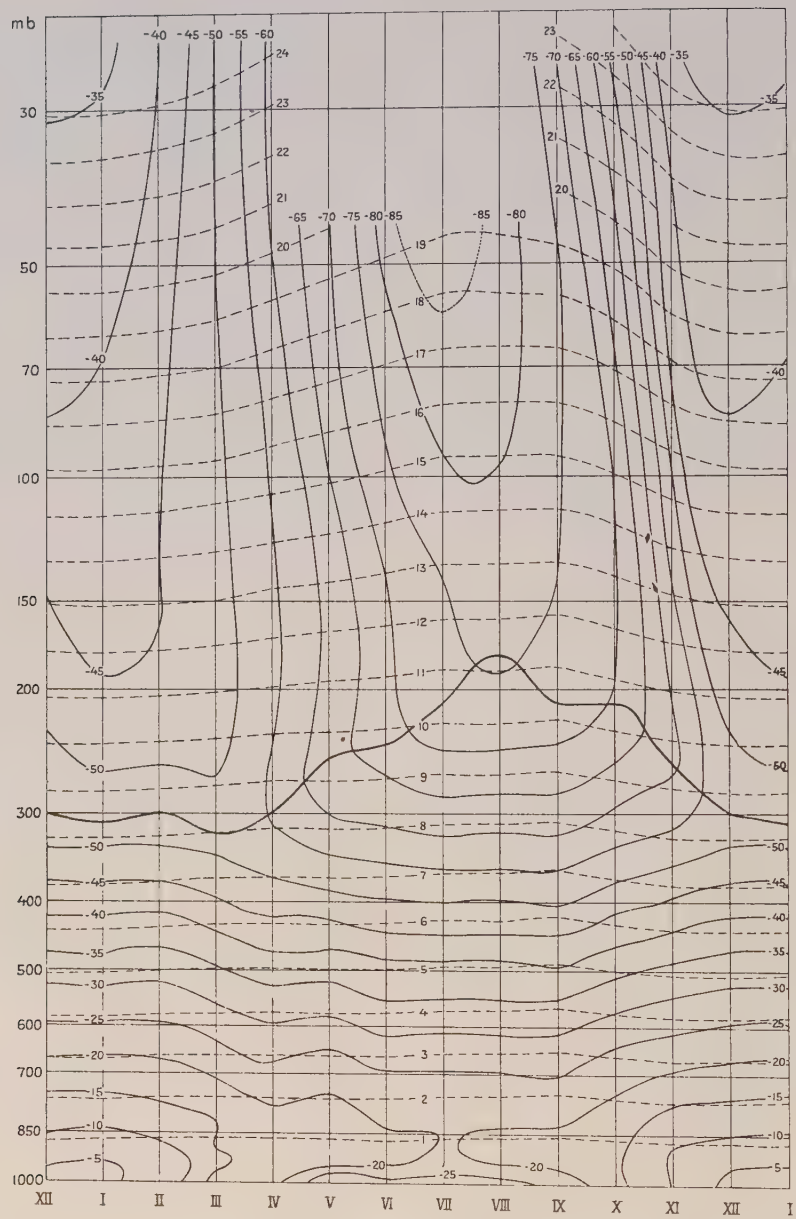


Fig. 2. Isopleth diagram. Full lines isotherms. Dashed lines heights in geopotential kilometres. Heavy full line the tropopause.

On the basis of the monthly mean curves for both years combined an isopleth diagram has been constructed, showing the yearly course of the mean temperature at any level as well as the course of the tropopause, as derived from the same curves (figure 2). The heights entered on the diagram have been calculated from the mean temperature values. For the highest levels the mean temperatures are based on the well-known method of mean differences, although different authors have raised strong objections against this method. (SCHERHAG, 1948). This question will be discussed more closely in the final expedition

report. Possible errors arising from the method used are of minor importance to the conclusions in the present note.

The troposphere. In the troposphere the mean temperature in summer decreases all the way from the ground up to the tropopause. Strictly speaking, this applies to day time conditions only, because a considerable daily variation is present in the lowest hundreds of metres. In the months of January and February the mean temperature near the surface was —4° to —5°C, that means considerably lower than in the corresponding months in the Arctic. Thus, over the north polar pack ice, SVERDRUP (1933) found in the warmest month a mean temperature of +3°C at 600 metres. Apart from the very highest layers, the whole troposphere is in summer considerably warmer over the Arctic than over the Antarctic. (Figure 3.)

During the autumn the troposphere is gradually cooled, especially in the lowest kilometre in connection with the formation of a surface inversion. In winter, when the inversion becomes persistent, there is in the lowest kilometre a mean temperature rise upwards of some 7°C. Thus, close to the ground a considerable annual temperature variation is found, although the range is somewhat smaller at Maudheim than at Arctic Bay. (Figure 3.)

Above the inversion layer the range rapidly decreases, and at 700 mb the difference between the highest and lowest monthly mean temperature is only 7°C, less than half the corresponding value from Arctic Bay.

This small annual temperature change of the middle troposphere seems to take place mainly in spring and autumn, whereas the month to month variation during summer and winter is extremely small. Thus, no distinct maximum and minimum can be pointed out, at least not from our comparatively short period of observation.

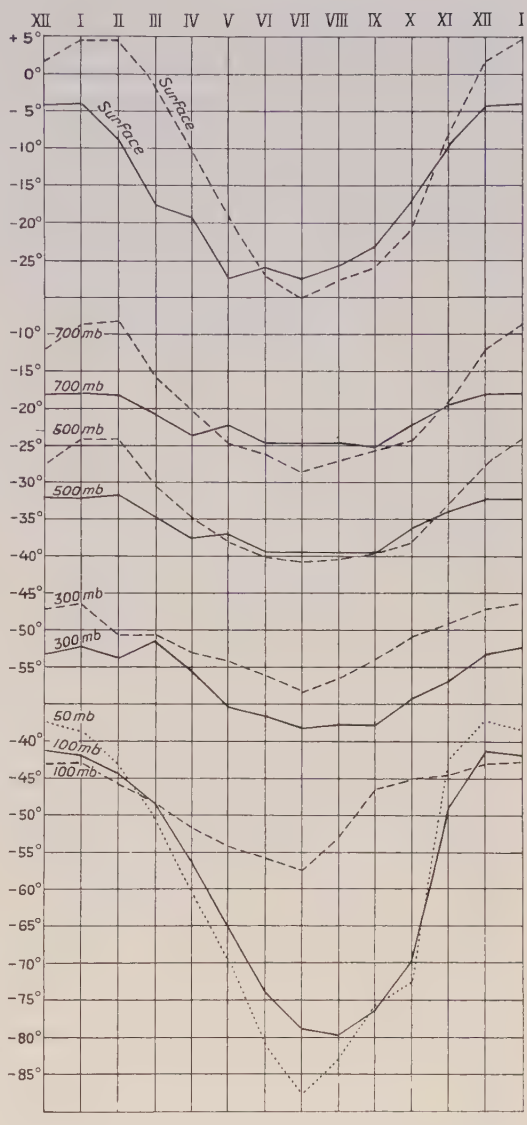


Fig. 3. Mean temperatures at some standard pressures for different months. Month 1 January and July for Maudheim and Arctic Bay respectively. Full lines Maudheim values. Dashed lines Arctic Bay values. Dotted line Maudheim temperatures at 50 mb.

The stratosphere. The great seasonal variation in the stratosphere is the most striking feature of the mean temperature distribution, as shown in figure 2. Starting with the highest levels, we find a distinct maximum in December, i.e. when the sun has its highest altitude, and the same applies to all the layers above the 100 mb surface. At this time of the year the lapse-rate is negative through the whole stratosphere, from the tropopause up to at least 25 mb. Soon after summer-solstice the highest layers start to cool off, at first slowly, but more rapidly in the early autumn. The stratification, therefore, becomes gradually more isothermal, and is almost exactly isothermal in March. As the

cooling continues and remains greatest in the uppermost layers, a temperature decrease with height above the tropopause becomes more and more pronounced during the autumn and the first part of the winter. At winter-solstice, or possibly a few weeks later, the stratospheric lapse-rate attains its highest positive value, exceeding 2°C per 1,000 metres from a height of 13 kilometres in July. In this month the lowest mean value at the 50 mb surface is found, the minimum being very distinct. In August, the warming of the highest layers starts, but in that month the temperature in the lowest part of the stratosphere reaches its minimum value.

This time lag in the occurence of the yearly minimum at different levels, produces a marked skewness in the diagram shown in figure 2. This marked feature in the temperature distribution is fairly well established, although in the coldest months the absolute temperature values are rather unreliable above 85 mb, due to very few readings. In July, all ascents above 100 mb showed a marked temperature decrease with height, but in August the stratification was more isothermal.

A heating of the highest layers is noticeable in late winter, but the main warming takes place after spring equinox. At 50 mb, the most rapid increase in temperature is found between October and November, or in the middle of October according to the half monthly mean values. From the first to the second half period of October, the temperature at that level rose from −69.5 to −57.4°C in the first year and from −71.4 to −51.0° C in the second, with a mean temperature rise of 16°C for both years combined, i.e. one degree per day.

Due to this pronounced heating of the highest layers in early spring, the stratospheric lapse-rate undergoes a rapid change in this period. In fact, the mean lapse-rate changes from its highest positive to its highest negative value in about three months. In the lowest part of the stratosphere the main heating takes place in late spring and early summer.

Comparing our results with arctic conditions, we find remarkable contrasts in winter, whereas the summer conditions are quite similar, as previously pointed out by COURT (1942). Unfortunately, aerological data from the Arctic are still scanty, especially for the higher layers in winter. To obtain the values at the highest level for Arctic Bay, figure 3, the mean ascent curve for the coldest months had to be extrapolated (HENRY and ARMSTRONG (1949)).

Figure 3 shows that in summer Maudheim is slightly warmer than Arctic Bay at 100 mb, whereas in the coldest months it is some 20 to 25°C colder. The yearly range at this level is 39°C and 15°C, respectively, the values thus differing by some 25° C. This difference probably increases with height.

Tropopause layer. Figure 2 demonstrates a great annual variation in the height of the tropopause over Maudheim. In the summer months, when the temperature of the lower stratosphere is comparatively constant, the tropopause is found near the 300 mb surface. In the autumn, its mean height increases as the temperature of the lower stratosphere decreases, reaching its highest value in winter when the temperature in the layer just above has its minimum. As this part of the stratosphere becomes warmer again in spring, the height of the tropopause gradually decreases, reaching "normal" summer value at the same time as the temperature of the 250 mb surface reaches its summer value.

It is of interest to examine if this intimate connection between the height of the tropopause and the stratospheric temperature means that the tropospheric conditions are without significance. For this purpose, the pressure at the tropopause (y) has been correlated with the temperature at the 600 mb (x_1) and the 150 mb (x_2) levels, all three values taken from each of the 43 half-monthly mean ascent curves.

Calling the correlation coefficient between y and x_1 r_{yx_1}, and that between y and x_2 r_{yx_2}, we found $r_{yx_1} = +0.48$ and $r_{yx_2} = +0.91$. The correlation coefficient between x_1 and x_2 was rather high, $r_{x_1x_2} = +0.74$, and the partial coefficients were found to have the following values $r_{x_1y \cdot x_2} = -0.73$ and $r_{x_2y \cdot x_1} = +0.95$.

These figures clearly show how the correlation coefficient r_{yx_1} gives a wrong impression of the connection between the height of the tropopause and the temperature of the lower troposphere. The linear regression equation will be:

$$y = 331 - 7.0x_1 + 4.5x_2$$

This shows that a given change in temperature at one of the two layers under consideration will be followed by the greatest change in the mean height of the tropopause if it occurs at the 600 mb level. However, the greater variation in the stratospheric mean temperature will make the last term of the equation dominate when seasonal changes are considered.

On the cause of the observed temperature distribution

In summer the troposphere is much colder in the Antarctic than in the Arctic. This striking difference is probably caused by the unequal values of the albedo in the two regions. At Maudheim, LILJEQUIST (1954) carried out extensive radiation measurements and found remarkably high albedo values even in summer. Corresponding measurements in the Arctic show much lower values, meaning that there the surface will receive more energy from the sun and accordingly attains higher summer temperatures. By various thermodynamical processes parts of this energy are transferred to the troposphere.

Before discussing the observed stratospheric temperatures, a few general remarks about the radiation in this part of the atmosphere are needed. According to DOBSON *et collab.* (1946) the main absorbing gases are carbon dioxide, water vapour and ozone. The amounts of water vapour as well as carbon dioxide generally decrease from the ground upwards, whereas ozone is found mainly in the stratosphere, the amount increasing with height in the layers under consideration.

To our discussion the absorption power of ozone in the infrared is of special interest. The main absorption band of the ozone lies at about 9.7 μ, that is in a part of the spectrum where both water vapour and carbon dioxide are nearly transparent. Therefore, most of the energy that is absorbed by the ozone comes from the surface or from the tops of the clouds, and depends particularly on the surface temperature (DOBSON 1946).

At Maudheim, the sun is below the horizon for approximately two months. On the other hand, according to a diagram by COURT (1942) the direct solar beams reach all the layers of the stratosphere for a short period during the day even in the darkest season. Still, the influence of the direct solar radiation may safely be neglected during this part of the year because of the great negative elevation of the sun and the short duration of sunshine. Furthermore, the absorption in the infrared by ozone must be of minor importance, as the surface temperature of the antarctic continent is extremely low. Due to strong circulation and slow adjustment to radiative equilibrium in the stratosphere in winter the observed temperature values at Maudheim can be expected to be highly dependent on the radiative conditions over the continent.

Thus, in the coldest months the temperatures of the lower stratosphere must be determined mainly by absorption and radiation in the infrared of carbon dioxide and water vapour. Dobson's estimates of the radiative equilibrium temperatures in the lower stratosphere of each of the two gases are 200° A. and 190° A., respectively, in remarkable agreement with the observed values. Furthermore, with decreasing amount upwards of the absorbing gases, the temperature has to decrease with height, again in accordance with our observations.

The heating of the highest layers which starts soon after winter solstice is most likely due to increasing intensity of the direct solar radiation. The flux of the long-wave radiation from below must remain comparatively unchanged, because at this time of the year no rise in temperature can be noticed in the troposphere or at the surface, and accordingly no heating of the stratosphere can arise from absorption of this radiation.

On the basis of these assumptions the skewness on the isopleth diagram (fig. 2) is easily explained. Firstly, the duration and intensity of sunshine increase with height, meaning that the radiation from the returning sun must be of greater importance at higher levels than further down. Secondly, because of the downwards increasing density of the air a greater amount of absorbed energy is needed to produce a noticeable temperature rise in the lowest part of the stratosphere. The relatively small amount of ozone in the lowest part of the stratosphere lends added importance to the last named factor. Altogether it is evident that a temperature rise caused by increasing sun radiation must be observed first at higher levels.

As the altitude of the sun increases in spring,

the warming of the stratosphere takes place with increasing rapidity. However, as the lowest part of the atmosphere including the surface simultaneously is heated, the long-wave radiation from below attains greater significance, especially to the conditions in the lowest parts of the stratosphere. Furthermore, if the ozone distribution shows variations that are similar to those observed in the northern hemisphere, the amount of ozone increases rapidly in late winter. There exist, therefore, valid reasons for assuming that the observed heating in spring as well as the rapid change in the lapse-rate are caused by the growing significance of ozone. Moreover, there are strong indications that at least above 100 mb, the absorption of the direct solar radiation is the dominating factor, even in spring and late summer. The rate of warming observed in October, exceeding 1°C day in some layers, is much greater than the maximum value to be expected from absorption of the outgoing radiation.

It should be added, that in the layers just above the tropopause, other factors, such as vertical motion extending from the troposphere, may be of considerable importance to the temperature distribution.

Next, we turn our attention to the observed difference between arctic and antarctic temperature in the stratosphere during winter. Court, who first discussed this marked contrast, concludes that a stronger meridional exchange of air in the northern hemisphere is the main reason for the higher temperature. Rubin, dealing with the same problems, shows that outgoing radiation may account for the observed cooling of the antarctic stratosphere from autumn to winter, and agrees with Court in his explanation of the higher values over the Arctic (Rubin, 1953). Both authors arrive at this conclusion, as they consider the available outgoing radiative energy from the surface too small to account for the difference.

When the upper wind data from Maudheim have been examined, more light will be thrown upon this matter. Therefore, a more detailed discussion will be presented in our final expedition report. At present, it will only be showed that there really exists radiative effect which may account for most, perhaps for all of the observed temperature difference.

A rough estimate can be made of the amount of energy needed to produce the difference. For comparison, the station Tromsø (69° 39′ N., 18°57′ E.) had to be used (Johansen, 1949), as no winter data for the highest layers were available from Arctic Bay.

Table II. Mean temperatures of the air mass between the 250 and the 50 mb surfaces

	I	II	III	IV	V	VI
Maudheim	− 43.6	− 45.0	− 47.6	− 54.6	− 63.9	− 71.4
Tromsø	− 45.0	− 46.9	− 49.9	− 53.8	− 59.3	− 59.7

	VII	VIII	IX	X	XI	XII
Maudheim	− 76.2	− 77.0	− 74.2	− 68.9	− 53.9	− 44.6
Tromsø	− 63.0	− 60.9	− 56.7	− 53.1	− 46.1	− 45.0

Table II shows mean temperatures in corresponding months for the air mass between the pressure surfaces 250 and 50 mb. The difference between the two figures for each month is proportional to the net amount o heat that is needed for raising the temperature from the lower to the higher value.

In month IV the temperatures at both stations are practically equal, but three months later there is a difference of some 13°C. This means that as compared to Maudheim, the air column over Tromsø has gained a net amount of heat of approximately 7 gcal. cm^{-2} day^{-1}. In addition, we must take into account the excess of lost energy due to radiation at higher temperatures. For this rough estimate we will assume the loss of heat by the air column under consideration to be equal to 20 % of the total black body radiation, as previously done by Elsasser in a similar connection (Elsasser, 1942). Using the mean temperature during the three months considered for both stations, we find an excess in lost energy at Tromsø of some 6 gcal. cm^{-2} day^{-1}. In other words, to produce the observed temperature contrast, the air mass between the 250 and 50 mb levels over Tromsø must have received an amount of heat of some 10—15 gcal. cm^{-2} day^{-1} in addition to the amount received by the same air column over Maudheim.

It is not unlikely that the absorption by ozone in the 9.7 μ band may produce a difference in absorbed energy of the same order of

magnitude. Over the antarctic continent, the surface temperatures in winter are extremely low. Temperature measurements in the ice made on the inland plateau by the glaciologist Charles Swithinbank on a weasel journey from Maudheim to the border of the inland ice cap indicated a mean surface temperature for the whole year of some – 40° C (Robin, 1953). Further inland it must be even colder. Thus, a mean temperature of some 220° A for the months under consideration does not appear unreasonable.

In the Arctic, the winter temperatures are much higher. In addition, the more extensive cloud cover may rise the effective temperatures for the outgoing radiation. With a temperature some 30° higher than the estimated value over the Antarctic, the intensity in the 9.5 μ band increases from 11 gcal. cm^{-2} day^{-1} to 24 gcal. cm^{-2} day^{-1} within a band width of 1 μ.

Dobson states that some 70—80 % of the energy is absorbed by the ozone, and if this absorption takes place in the lower part of the polar stratosphere in winter, it seems to be sufficient to account for the observed temperature difference. This effect may also explain the more isothermal stratification observed in the arctic stratosphere at this time of the year.

REFERENCES

Court, A., 1942: Tropopause disappearance during the antarctic winter. *Bull. Amer. meteor. Soc.*, **23**, pp. 220—238.

Dobson, G. M. B., Brewer, A. W., and Cwilong, B. M., 1946: Meteorology of the lower stratosphere. (Bakerian Lecture) *Proc. Roy. Soc. London, A*, **185**, pp. 144—175.

Elsasser, W. M., 1942: Heat transfer by infrared radiation in the atmosphere. *Harvard meteor. Stud.*, No. 6, p. 97.

Gibbs, W. J., Gotley, A. V., and Martin, A. R., 1950: Heard and Macquary Islands, 1948. *Austral. natl. Antarctic Res. Exped. Rep.*, Ser. D., **1**, part 1 a, p. iii.

Henry, T. J. G., and G. R. Armstrong, 1949: Aerological data for northern Canada. Toronto, Dept. of Transport, pp. 132—145.

Johansen, H., 1949: Mean upper air data obtained from soundings at Tromsø during the years 1941—44. *Geofys. Publ.*, **17**, No. 4, pp. 16—17.

Liljequist, G. H., 1954: Radiation, and wind and temperature profiles over an Antarctic snow-field. A preliminary note. *Proceedings of the Toronto Meteorological Conference* 1953.

Raunio, N., 1951: Amendments to the computation of the radiation error of the Finnish radiosonde. *Geophysica*, **4**, pp. 14—19.

Robin, G., 1952: Queen Maud Land. The scientific results of the international expedition. *The Geographical Magazine*, p. 289.

Rubin, M. J., 1953: Seasonal variations of the Antarctic tropopause. *J. Meteor.*, **10**, 127—134.

Scherhag, R., 1948: *Wetteranalyse und Wetterprognose*, Berlin, Springer-Verlag, p. 59.

Sverdrup, H. U., 1933: Meteorology, The Norwegian North Polar Expedition with the Maud, 1919—25. II, Pt. I, Discussion, pp. 38—39.

Väisälä, V., 1941: Der Strahlungsfehler der finnischen Radiosonde. Univ. of Helsinki, Inst. of Meteor. Mitteilungen No. 47.

Die Entstehung einer Höhenzyklone über Nordamerika[1]

Von E. KLEINSCHMIDT, Max Planck-Institut für Strömungsforschung, Göttingen

(Manuscript received April 5, 1954. Revised manuscript received November 4, 1954.)

Abstract

The existence of a high-level cyclone always depends on masses of increased potential vorticity in the upper troposphere. The origin of those masses is studied in a single case of cut-off. It is found that the upper part of them comes from the polar stratosphere, while the lower part consists of tropospheric air. Both from the vorticity equation and from the single wind observations, the existence of a non-conservative force results, parallel and south to the jet-stream. This force can be regarded as the immediate cause of the cut-off. An explanation of the force is not attempted.

Einleitung

Den Anlaß zu der folgenden Untersuchung gab eine scheinbar sehr spezielle Frage. Es ist die Frage nach der Entstehung des sogenannten *Tropopausen-Trichters*. In jeder hochreichenden Zyklone findet man jene Massen, die ihrer potentiellen Temperatur nach der oberen Troposphäre angehören, die aber dabei so stabil geschichtet sind, daß man sie oft zur Stratosphäre rechnen möchte. Als die Messungen aus der hohen Atmosphäre noch spärlich waren, hat man ihre Untergrenze oft als trichterförmige Einsenkung der Tropopause aufgefaßt. Bei genügend dichten Beobachtungen zeigt sich aber, daß die Begrenzung nie von dieser einfachen Form ist und daß die Massen des »Trichters« immer mindestens auf einer Seite stetig in die umgebende Troposphäre übergehen. Trotzdem wollen wir an der Bezeichnung Trichter festhalten.

Daß man derartige Massen immer nur in Höhenzyklonen findet, ist kein Zufall, sondern einfach eine Folge des Gradientwind-Gleichgewichts. Dort, wo die vertikale Schichtung stabiler ist als in der horizontalen Umgebung, muß im Gleichgewicht ein relatives (oder absolutes) Maximum an zyklonaler Rotation gegenüber den darunter und darüber liegenden Schichten vorhanden sein. Da dieser Zusammenhang gegenseitig ist, gehören Tropopausen-Trichter und Höhenzyklone untrennbar zusammen.

Damit ist aber noch nicht alles gesagt. Vielmehr ist es oft so, daß auf den Massen des Tropopausen-Trichters allein die Existenz der ganzen Zyklone beruht. Denn sie allein besitzen den gegenüber der horizontalen Umgebung überhöhten potentiellen Wirbelwert, ohne den, wie der Verfasser (KLEINSCHMIDT, 1950) gezeigt hat, keine Zyklone existieren

[1] Die Arbeit wurde 1952 anläßlich eines halbjährigen Aufenthalts am Meteorologischen Institut der Universität Stockholm begonnen und im Max Planck-Institut für Strömungsforschung fortgesetzt. Sie wurde von der Math.-Naturwissenschaftlichen Fakultät der Universität Göttingen als Habilitationsschrift angenommen.

kann. Im Tropopausen-Trichter erreicht der potentielle Wirbelwert manchmal das 8-fache von dem der umgebenden Troposphäre. In den höheren und tieferen Schichten herrschen dagegen im allgemeinen nur geringe oder gar keine Gegensätze.

Daraus folgt, daß das Problem der Höhen-Zyklogenese im Wesentlichen in der Erzeugung jenes hohen potentiellen Wirbelwerts liegt. Daß er schon vor der Zyklone vorhanden war, ist nicht möglich, jedenfalls nicht bei derselben Anordnung der Massen. Denn auch der Zusammenhang zwischen potentiellem Wirbelwert und Zyklone ist gegenseitig. Eine Zyklone verlangt einen überhöhten Wirbelwert, ein überhöhter Wirbelwert verlangt aber auch eine Zyklone. Folglich muß jene charakteristische Anordnung der potentiellen Wirbelwerte genau gleichzeitig mit der Zyklone entstanden sein.

Es ist sehr unwahrscheinlich, daß die Massen des Trichters ursprünglich den normalen troposphärischen Wirbelwert besessen haben. Nach dem Satz von Ertel (1942) sind zu jeder Änderung des potentiellen Wirbelwerts entweder Wärme- und Kältequellen oder nichtkonservative, zusätzliche, d. h. nicht in der Eulerschen Gleichung enthaltene Kräfte erforderlich. Solche Kräfte sind die auf Reibung und Austausch beruhenden Kräfte. Nun sieht man aber in den Höhenwetterkarten die Zyklonen oft in kurzer Zeit entstehen. Wodurch man sich das Anwachsen des potentiellen Wirbelwerts auch hervorgerufen denkt, in jedem Fall müßten jene Einflüsse überaus stark sein, viel stärker, als man nach Theorie und Erfahrung anzunehmen geneigt ist.

Aus demselben Grund ist es unwahrscheinlich, daß der Tropopausen-Trichter ein abgesunkenes Stück Stratosphäre ist. Dann brauchte sich zwar der potentielle Wirbelwert nicht zu ändern. Da aber das Stück innerhalb der kurzen Zeit der Zyklogenese absinken müßte, wäre eine abnorm starke Abkühlung erforderlich.

Die beste Erklärung für die Entstehung des Tropopausen-Trichters scheint zu sein, daß seine Massen aus der polaren Stratosphäre stammen, die nördlich des Westwindbandes beginnt. Dort liegen ausgedehnte Massen von etwa der richtigen potentiellen Temperatur und dem richtigen potentiellen Wirbelwert. Die Entstehung der Höhenzyklonen hätte man sich dann so vorzustellen, daß eine Teilmasse aus den unteren Schichten der polaren Stratosphäre herausgelöst und in die Westdrift oder auf deren Südseite versetzt wird. Für die Richtigkeit dieser Annahme spricht, daß die Höhenzyklonen fast stets im Bereich des Westwindbandes als cut-off-Zyklonen entstehen. Allerdings kann man sich nur schwer vorstellen wodurch die Abtrennung der Teilmasse bewirkt werden soll. Wahrscheinlich sind auch hierzu erhebliche »zusätzliche« Kräfte erforderlich.

In jedem Fall ist die Entstehung eines Tropopausen-Trichters ein problematischer Vorgang, bei dem das im allgemeinen gültige Gesetz nahezu reibungslosen und adiabatischen Verhaltens in irgend einer Weise durchbrochen zu sein scheint.

Um hier wenigstens in einem Einzelfall Klarheit zu gewinnen, wurde eine Zyklone ausgewählt, die im entwickelten Zustand einen gut ausgeprägten Tropopausen-Trichter besaß, und deren Bildung sich innerhalb kurzer Zeit über Nordamerika, also einem gut mit Radiosonden versehenen Gebiet abspielte. Wenn überhaupt, so müßte in einem solchen Fall eine Antwort auf die angeschnittenen Fragen zu erhalten sein.

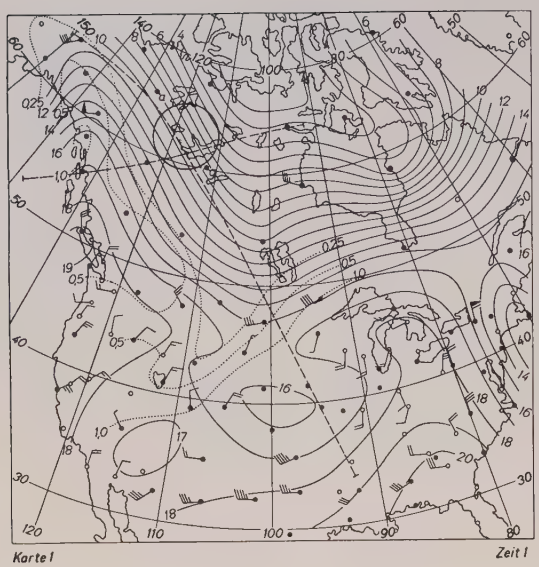
Die Zyklone vom 7. April 1951

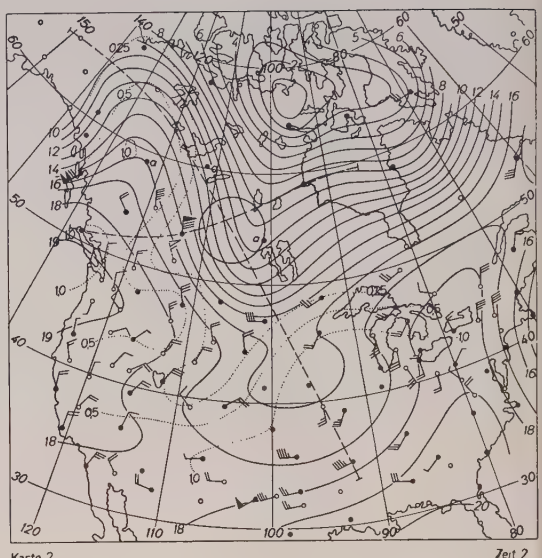
Die Karten 1—4 zeigen das Höhenstromfeld in 12-stündigem Abstand. Sie umfassen den Zeitraum vom 6. April 03^h GMT (Zeit 1) bis zum 7. April 15^h GMT (Zeit 4). Das Westwindband hatte seit 4—5 Tagen ziemlich unverändert über Kanada gelegen, bis es plötzlich begann, sich nach Süden auszubuchten. Aus der Spitze des immer schärfer werdenden Trogs entwickelte sich die neue Höhenzyklone.

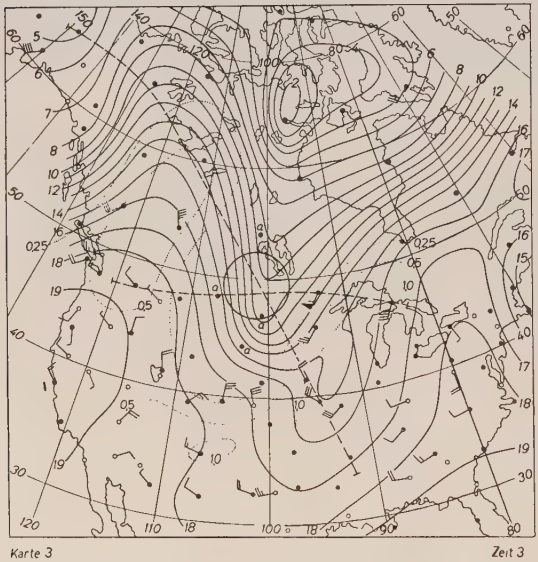
Die Karten geben das Stromfeld in der isentropen Fläche $\vartheta = 35°$. Sie liegt auf der polaren Seite des Westwindbands in etwa 9, auf der subtropischen Seite in etwa 5 km Höhe. Die Darstellung in Isentropenkarten wurde statt der üblichen in Isobarkarten gewählt, weil mit starken vertikalen Versetzungen zu rechnen ist. Wenn auch vielleicht Wärmequellen vorhanden sind, so sind doch die Luftteilchen vermutlich immer noch eher an die isentropen als an die isobaren Flächen gebunden.

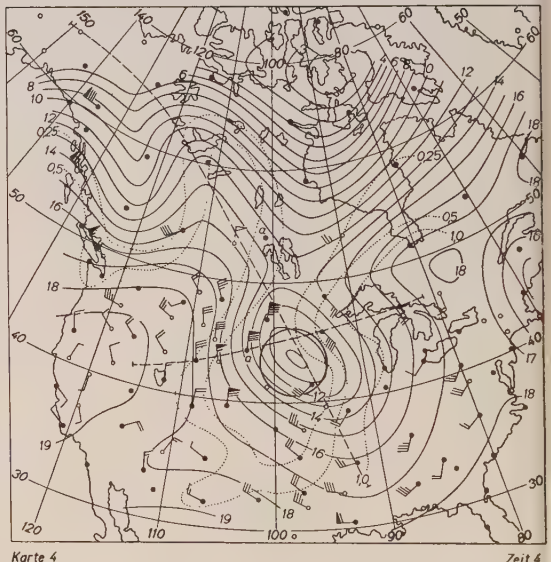
In die Karten sind außer den gemessenen Winden die Linien $\Psi = \text{const}$ eingetragen, wo

$$\Psi \equiv c_p T + \Phi$$

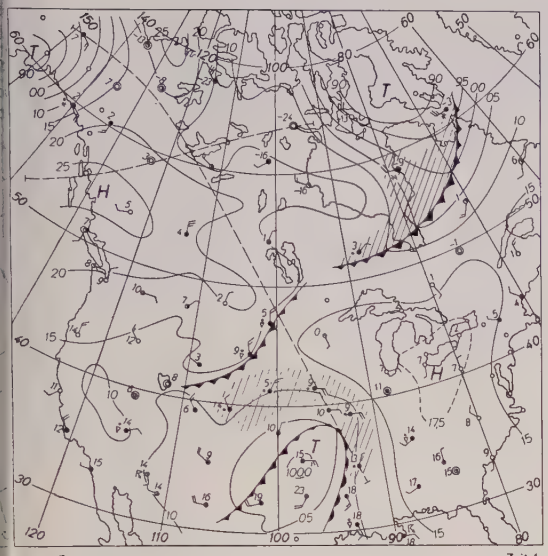


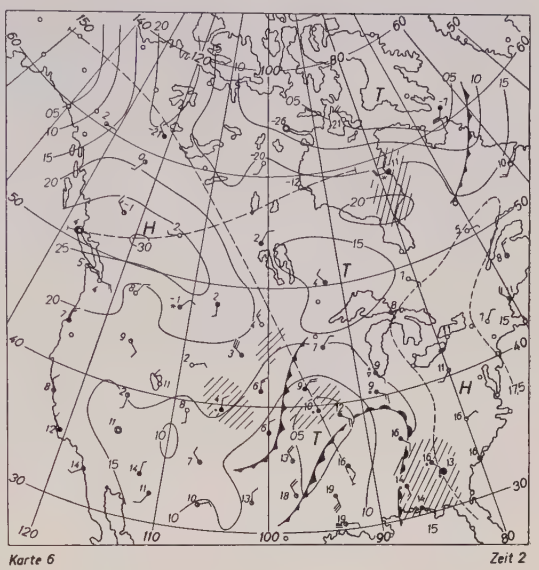


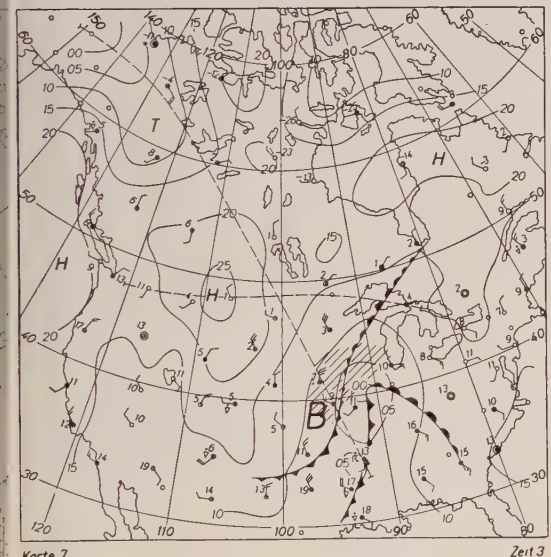


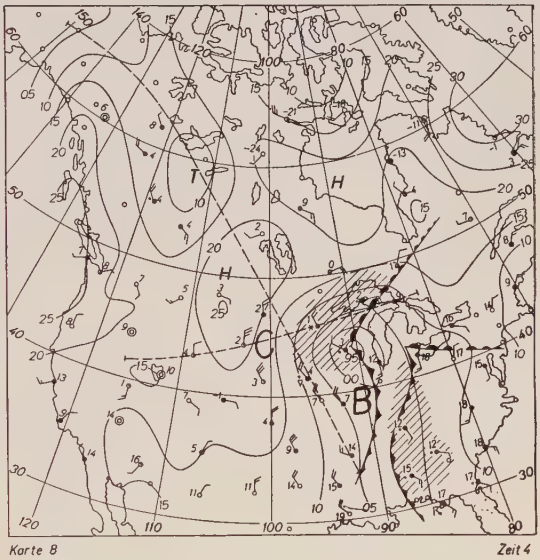


Karten 1—4: Isentropenkarten für die Fläche $\vartheta = 35°$. Ausgezogen: Linien $\vartheta = \text{const}$. Gestrichelt: Spurlinien der Vertikalschnitte. Punktiert: Linien gleicher spez. Feuchte.









Karten 5—8: Bodenwetterkarten. Gestrichelt: Spurlinien der Vertikalschnitte.

die auf die Masseneinheit bezogene Summe aus Enthalpie und Geopotential ist. Wie MONTGOMERY (1937) gezeigt hat, spielt Ψ auf den isentropen Flächen eine ähnliche Rolle wie Φ auf den isobaren. Die Linien $\Psi = \text{const}$ sind Stromlinien des geostrophischen Windes. Seine Geschwindigkeit beträgt

$$v_g = \frac{|\text{Grad } \Psi|}{f},$$

wo f der Coriolisparameter und Grad der in den Isentropenkarten gebildete Gradient ist. Der Wert von Ψ wurde für die schwarz ausgefüllten Stationen berechnet, und danach die Linien von 400 zu 400 m^2sec^{-2} gezeichnet. Die Linien sind laufend numeriert, die Linie Null entspricht dem niedrigsten auf den Karten erreichten Wert von Ψ.

Die Karten 5—8 sind die Bodenwetterkarten für die Zeiten 1—4. Sie zeigen anfangs eine schwache Zyklone mit breiter Warmluftzunge, die langsam von Texas nach NE zieht und dabei allmählich erlischt. Gleichzeitig dringt von NW eine instabil geschichtete kühlere Luftmasse vor. An ihrer Vordergrenze entwickelt sich von Zeit 2 an ein Gebiet kräftiger Niederschläge und eine stark zyklonale Bodenströmung. Es scheint, daß es sich hierbei um den Vorgang handelt, den der Verfasser (KLEINSCHMIDT, 1950) beschrieben hat, nämlich um die Erhöhung des potentiellen Wirbelwerts in Bodennähe durch die freiwerdende Kondensationswärme. Daß diese Vermutung richtig ist, bestätigt der vertikale Aufbau der neuen Zyklone.

Den Aufbau der Zyklone zur Zeit 4 zeigen die beiden Vertikalschnitte 4 und 8. Sie schneiden sich ungefähr rechtwinklig längs der Achse der Höhenzyklone. Ihre Spurlinien sind in den Karten 4 und 8 gestrichelt eingetragen. Die auffallendste Erscheinung in den beiden Schnitten ist die Masse A, der Tropopausen-Trichter. Aus den Windmessungen in der 35°-Fläche ergibt sich in seinem Innern eine maximale Rotation von $17 \cdot 10^{-5}$ sec^{-1}, bei einem f von $10 \cdot 10^{-5}$ also eine Absolutrotation von $27 \cdot 10^{-5}$ sec^{-1}. Da die Stabilität im Kern der Masse etwa doppelt so groß ist wie in der Umgebung, so erreicht hier der potentielle Wirbelwert das 5—6 fache des Außenwerts.

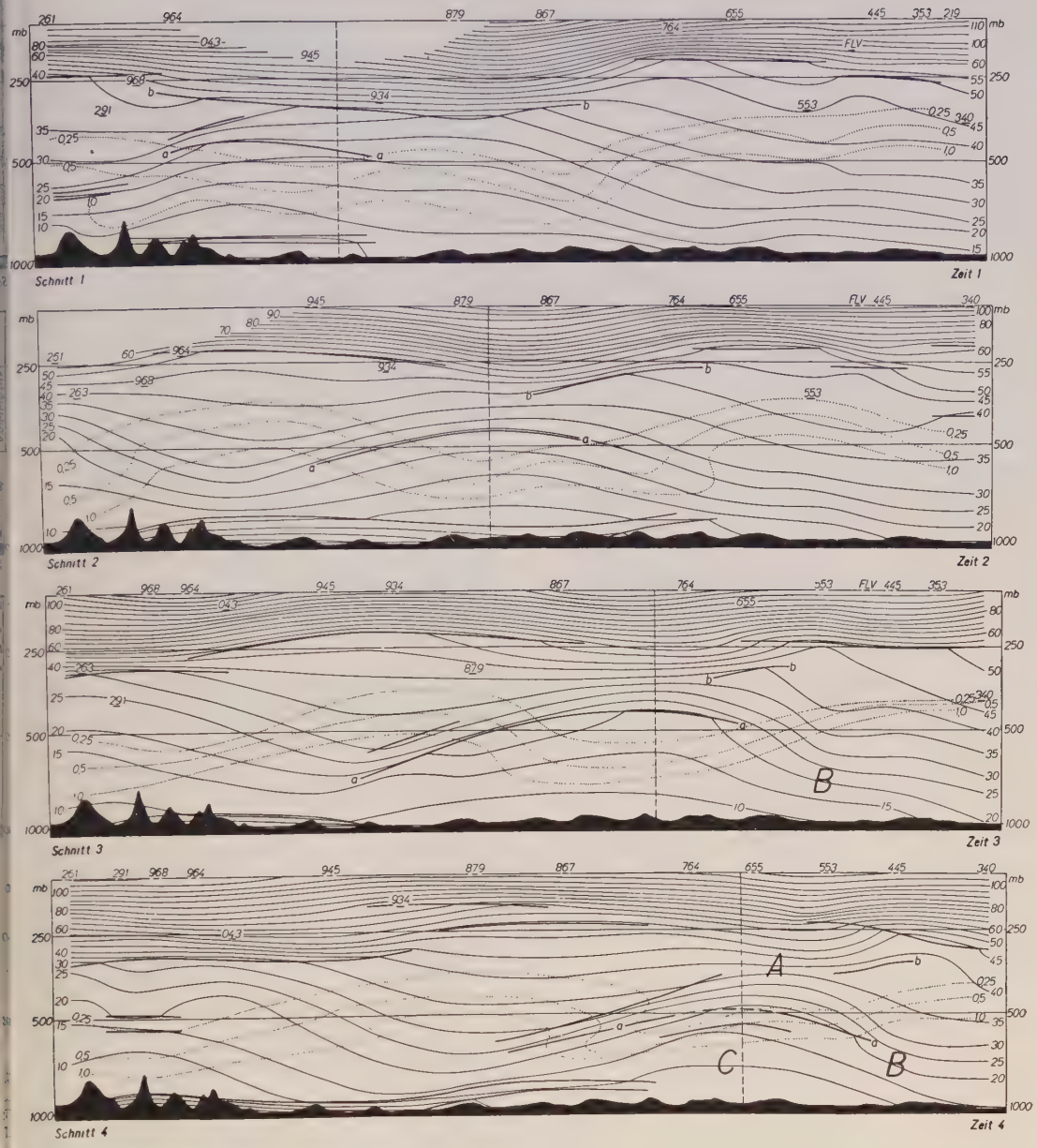
Der von dem Regengebiet erzeugte hohe potentielle Wirbelwert müßte in Bodennähe und etwa dort zu finden sein, wo in den Karten 7 und 8 der Buchstabe B eingetragen ist. Tatsächlich ist er auch in den Radiosonden von Omaha, Zeit 3 (553 in Schnitt 3) und von Columbia, Zeit 4 (445 in Schnitt 4) an der stabilen Schichtung zwischen 15°- und 30°-Fläche deutlich zu erkennen.

Auch die weitere Entwicklung spricht dafür, daß die zyklonale Rotation in Bodennähe durch die Kondensation hervorgerufen wurde. In den 24 Stunden nach Zeit 4 löst sich das Niederschlagsgebiet völlig auf. Sofort zerfällt auch die Bodenzyklone, obwohl sich die zyklonale Rotation in der Höhe noch verstärkt. Sowie die Produktion neuer Massen mit überhöhtem potentiellem Wirbelwert aufhört, gleicht die Reibung am Erdboden die bestehenden Unterschiede rasch aus.

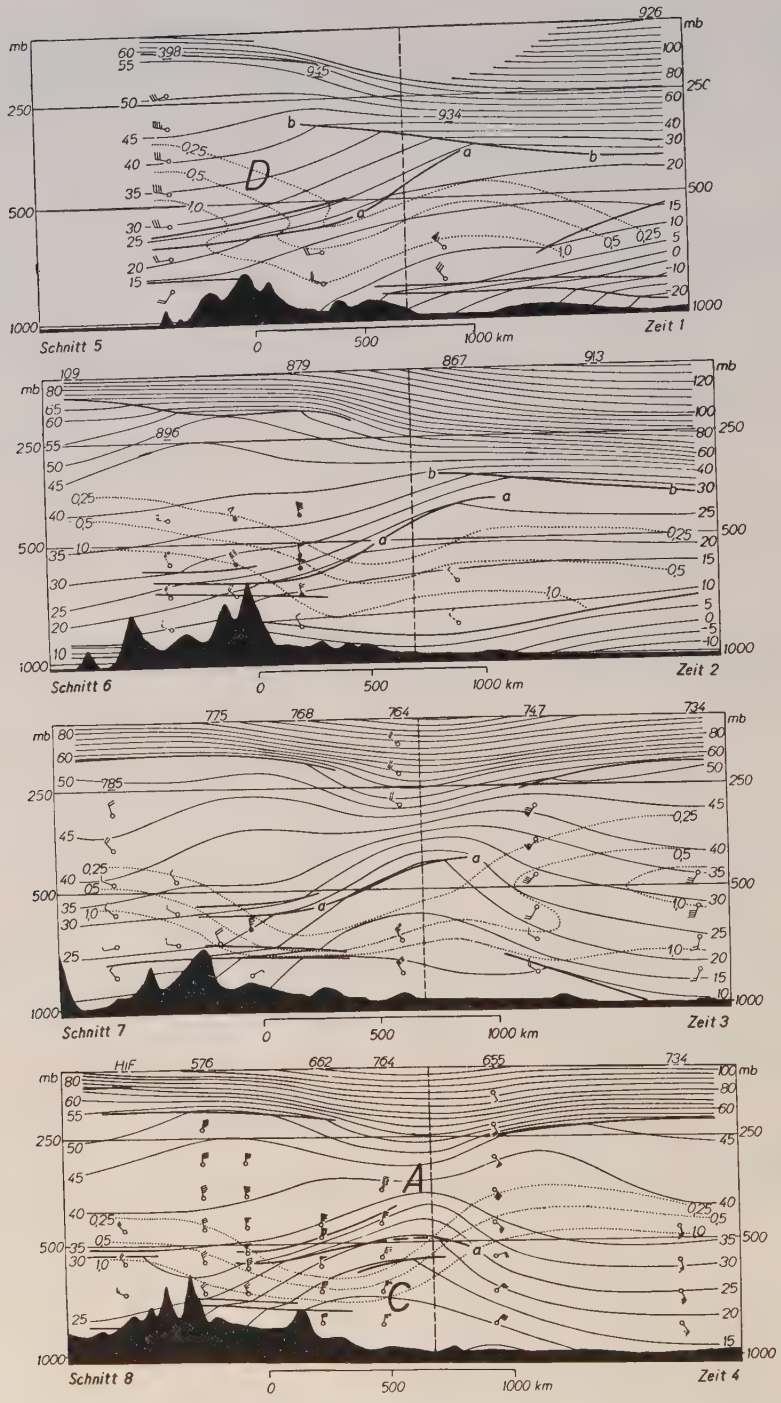
Auf dem hohen potentiellen Wirbelwert der Massen A und B beruht zur Zeit 4 die Zyklone in ihrer Gesamtheit. Die obere Masse liegt dabei weiter nordwestlich als die untere. Dem entspricht auch der Verlauf der Zyklonenachse. Oberhalb der 500 mb-Fläche steht sie, wie die Topographien der höheren Druckflächen zeigen, nahezu senkrecht. Unterhalb verläuft sie schräg bis ins Zentrum der Bodenzyklone.

Unmittelbar unter der Masse A liegt ein Berg potentiell kalter Luft (C in Schnitt 4 und 8). Diese Luftmasse gehörte noch 24 Stunden vorher zu der polaren bodennahen Kaltluft, die damals den größten Teil Kanadas bedeckte. Sie wurde mit der Nordströmung, die das neue Kondensationsgebiet auf seiner Westseite hervorrief, nach Süden befördert. Zur Zeit 4 hängt die Masse vielleicht noch über den Winnipeg-See mit der Muttermasse zusammen 12 Stunden später ist die Verbindung sicherlich gelöst.

Daß die Masse C einen Berg bildet und nicht auseinanderfließt, verdankt sie allein der Masse A. Eine Luftmasse mit übernormalem potentiellen Wirbelwert hebt im Gradientwindgleichgewicht die unter ihr liegenden Schichten an. Wahrscheinlich hat dieser Effekt auch die Niederschläge an der Kaltfront aufgelöst, die dann zur Bildung der Masse C führten. Wenn sich eine Luftmasse wie die Masse A relativ zu den tieferen Schichten bewegt, so treten Niederschläge immer unter ihrer Vorderseite auf, in der Richtung der Relativbewegung gesehen. Hier werden die unteren Massen angehoben, auf der Rückseite sinken



Vertikalschnitte 1—8: Dünn ausgezogen: isentrope Linien. Dick ausgezogen: Unstetigkeitsflächen in grad ϑ. Punktiert: Linien gleicher spezifischer Feuchte. Vertikale gestrichelte Linien: Schnittlinien mit jeweils dem umseitigen Vertikalschnitt.


mb
926
398
935
934
D
a
b
0,25
0,5
1,0
Schnitt 5
Zeit 1
0
500
1000 km
879
867
913
896
Schnitt 6
Zeit 2
775
768
764
747
734
785
Schnitt 7
Zeit 3
HIF
576
662
764
655
734
A
C
Schnitt 8
Zeit 4

sie wieder abwärts. So gesehen, ist das Niederschlagsgebiet, und folglich auch die Bodenzyklone, eine sekundäre Erscheinung, hervorgerufen durch das Herannahen der Masse A.

Ebenso, wie die Masse A die unter ihr liegenden Schichten anhebt, zieht sie die über ihr liegenden herab. Das zeigt sich in der (schwachen) Einsenkung der isentropen Flächen oberhalb von A. Wenn man den potentiellen Wirbelwert dieser hohen Schichten berechnet – wobei man in Ermangelung von Windmessungen auf den Gradientwind angewiesen ist, den man aus der Topographie der 200 mb-Fläche bestimmt –, so zeigt sich, daß er innerhalb der Zyklone etwa ebenso groß ist wie außerhalb, soweit sich das bei der großen Unsicherheit in jenen Topographien sagen läßt. Das läßt vermuten, daß bei unserer Zyklogenese in der Stratosphäre keine wesentlichen Wärmequellen und auch keine zusätzlichen Kräfte wirksam waren, daß die zyklonale Rotation hier oben allein durch die von der Masse A bewirkte vertikale Streckung entstanden ist.

Die Herkunft der Masse A

Es gilt, die Masse A rückwärts zu verfolgen. Da hohe Windmessungen nur spärlich vorhanden sind, muß der geostrophische Wind den ersten Anhalt für die Bahn der Masse liefern. Schnitt 4 wurde so gelegt, dass er die Bahn zwischen Zeit 1 und Zeit 4 enthält, falls die Masse in dieser Zeit sich adiabatisch verhalten hat und mit dem geostrophischen Wind gezogen ist. Die Schnitte 1—3, jeweils für die gleichnamige Zeit gültig, liegen längs derselben Spur. Jeder von ihnen müßte bei adiabatischem Verhalten und geostrophischer Versetzung die Masse A enthalten.

Da die Spur dieser »Längsschnitte« im allgemeinen nicht durch die Radiosondestellen geht, wurden für jeden Längsschnitt zuerst 8 querverlaufende Schnitte entworfen und diesen die Daten für den Längsschnitt entnommen. Auf diese Weise wurden sämtliche in einem breiten Streifen rechts und links der Bahn liegenden Beobachtungen verwertet. Von den »Querschnitten« ist in den Schnitten 5—8 jeweils derjenige wiedergegeben, in dessen Nähe die Masse bei geostrophischer Versetzung liegen müßte. Die Spurlinien sämtlicher Schnitte sind in den zeitlich entsprechenden Karten eingetragen.

Das *Ergebnis:* Zur Zeit 3 ist die Masse in Längs- wie im Querschnitt deutlich zu erkennen. Aber schon 12 Stunden weiter zurück wird ihre Lage in der Querrichtung unsicher. Im Schnitt 6 erscheint rechts die tiefe polare Stratosphäre. Eine klare Trennung zwischen ihr und der Masse A ist nicht mehr möglich.

Damit wird die Lage kritisch. Denn man kommt natürlich zu ganz verschiedenen Ergebnissen, je nachdem man den Standort der Masse im Schnitt 6 ein wenig mehr rechts oder links annimmt. Wenn die Zugrichtung zwischen Zeit 2 und Zeit 3 nur um 20° von der geostrophischen Windrichtung nach rechts abwich, so gehörte die Masse zur Zeit 2, immer adiabatisches Verhalten vorausgesetzt, schon völlig der polaren Stratosphäre an. Die nächstgelegenen Höhenwindmessungen sprechen für eine solche Abweichung, vor allem die Winde von Edmonton und Grande Prairie, in Karte 2 unmittelbar westlich des großen Kreises gelegen. Andererseits liegen diese Winde so weit entfernt vom Schwerpunkt der Masse, dass sie nicht unbedingt dessen Bewegung repräsentieren müssen.

Tatsächlich zeigt eine genauere Analyse, daß der Schwerpunkt doch annähernd dem Verlauf des Längsschnittes gefolgt ist. Die Masse A wird zur Zeit 3 nach unten durch eine scharfe Unstetigkeitsfläche begrenzt (Linien aa in Schnitt 3 und 7). Die 4 Radiosonden, die die Fläche belegen, sind in Karte 3 ebenfalls mit a bezeichnet. Nach Westen fällt die Fläche ab und geht in die Untergrenze einer stabilen Schicht von geringer vertikaler Mächtigkeit über. 12 Stunden früher, zur Zeit 2, ist dieselbe Fläche noch an 3 Punkten belegt (Stationen a in Karte 2). Am südlichsten (The Pas, 867) ist sie allerdings ziemlich diffus. Dafür zeigt dieser Aufstieg wesentlich höher, bei 327 mb, noch einen zweiten deutlichen Knick in der Temperaturkurve. Dieser Knick wurde als Schnittpunkt der tiefen polaren Tropopause (b in Schnitt 6) gedeutet. Daß diese Auffassung richtig ist, bestätigt sich weitere 12 Stunden vorher. Zur Zeit 1 springt die polare Tropopause als breiter Schirm nach S und SW vor (b in Schnitt 1 und 5), belegt von sämtlichen Stationen nördlich der punktierten Linien in Karte 1. Auch die Fläche a ist zu dieser Zeit noch an zwei Stationen vorhanden, sie hat aber anscheinend an Längsausdehnung verloren. Es zeigt sich also, daß die Masse A nicht von einheitli-

chem Ursprung ist. Der obere Teil ist stratosphärischer Herkunft, der untere gehörte zur Zeit 1 einer ziemlich stabilen troposphärischen Masse (D in Schnitt 5) an.

Weiter zurück haben sich, wie die früheren Radiosonden zeigen, die Luftmassen nicht mehr wesentlich verändert. Der vorspringende Schirm der polaren Tropopause wird stets in etwa derselben Höhe wiedergefunden. 24 Stunden vor Zeit 1 bedeckt er gerade Alaska. Um diese Zeit beginnt die stabile Masse D die Südküste Alaskas zu überschreiten. Ihre untere Begrenzung, zwischen 600 und 500 mb, ist gut ausgeprägt, ihre Vordergrenze dagegen ziemlich diffus, die Grenzfläche a ist als solche nicht mehr vorhanden. Sie hat sich wahrscheinlich erst kurz vor Zeit 1 gebildet.

Wenn noch ein Zweifel and der Richtigkeit des hier gezeichneten Bildes bestehen sollte, so wird er durch die Betrachtung des Feuchtefeldes behoben. In den Karten 1—4 sowie in allen Vertikalschnitten sind punktiert Linien konstanter spezifischer Feuchte q eingezeichnet, und zwar die Schnittlinien der Flächen $q = 1{,}0$, $q = 0{,}5$ und $q = 0{,}25$ gr/kg. Oberhalb der letzten dieser Flächen werden die Feuchtemessungen so spärlich, und außerdem so ungenau, daß sich hier kein sicheres Bild des Feuchtefeldes mehr entwerfen läßt. Doch ist das in unserem Fall auch nicht nötig. Wie aus den Schnitten zu erkennen, liegt jeweils zwischen der Fläche a und der Fläche $q = 0{,}25$ höchstens ein Abstand von 100 mb. Die polare Tropopause dagegen ist zur Zeit 1 und zu den früheren Zeiten durchweg mindestens 150 mb von jener Feuchtefläche entfernt. Es ist also ausgeschlossen, daß a etwa ein abgesunkenes Stück Tropopause ist.

Da die Fläche a zur Zeit 1 schräg auf die Tropopause stößt, muß der untere Teil der Masse A im Winkelraum zwischen beiden Flächen liegen. Er ist also wirklich ungefähr dem Längsschnitt gefolgt. Der Ort der oberen Teilmasse ist zwar nicht genau zu bestimmen, doch dürfte ihre Bahn nicht wesentlich von der der unteren abweichen.

In den Karten 1—4 sind die Standorte von A, wie sie sich aus den Schnitten ergeben, durch Kreise bezeichnet. Sehr genau ist die Ortsbestimmung natürlich nicht. Immerhin läßt sie erkennen, daß die Masse anfangs schneller (mit etwa 50 Knoten) gezogen ist als gegen Ende des Zeitraums (mit etwa 30 Knoten).

Wärme- und Kältequellen im Bereich der Masse A.

Nachdem nun die Bahn der Masse festliegt, gehen wir daran, ihr thermodynamisches und dynamisches Verhalten zu untersuchen. Ob in ihrem Bereich nennenswerte Wärmequellen liegen, läßt sich zwar im allgemeinen auch jetzt noch nicht feststellen. Nur an der Untergrenze bietet sich eine Möglichkeit. Soeben wurde schon erwähnt, daß sich die Fläche a innerhalb eines gewissen Abstandes von der obersten Feuchtefläche hält. Tatsächlich liegt dieser Abstand bei allen Radiosonden, die zwischen Zeit 1 und 3 die Fläche a in der Nähe ihres Scheitelpunkts durchstossen (es sind insgesamt 4), zwischen 70 und 90 mb. Das berechtigt zu der Annahme, daß die Fläche, mindestens in ihrem hochliegenden Teil, *massenfest* ist, so daß man aus den Änderungen der potentiellen Temperatur auf ihr Wärmezufuhr und -entzug berechnen kann.

Zur Zeit 1 sind es nur zwei Radiosonden, die die Fläche erfassen, die nördliche bei einer potentiellen Temperatur von 25°, die südliche bei einer solchen von 22°. Die drei Sonden von Zeit 2 melden dagegen alle 25—26°. Man kann daraus auf eine Erwärmung im ersten Zeitabschnitt schliessen, doch ist der Schluß nicht sehr überzeugend. Umso deutlicher ist die Abkühlung im nächsten Zeitintervall. Die zur Zeit 3 im Bereich der Masse A liegenden Stationen (3 an der Zahl) melden alle an der Fläche eine potentielle Temperatur von 20—21°. Die Fläche hat sich also um 5° potentiell abgekühlt. Dem entspricht nach der Formel

$$\triangle_p T = \frac{T}{\vartheta_{\text{abs.}}} \triangle \vartheta$$

eine Temperaturabnahme bei konstantem Druck ($\triangle_p T$) von 4°.

Einen Grund für diese Abkühlung, an deren Realität nicht zu zweifeln ist, kann der Verfasser nicht angeben. Die langwellige Ausstrahlung führt, soweit bekannt, nicht zu derart hohen Werten. Wolken, die als Ausstrahlungsflächen dienen könnten, sind bei der geringen Feuchte nicht vorhanden. Über Dunstobergrenzen ist nichts bekannt. Bestimmt man aus der vertikalen Feuchteverteilung im Bereich der Masse A nach dem von Elsasser (1942) berechneten Diagramm die langwelligen Strahlungsströme, so erhält man für $\triangle_p T$ Werte von

der Grössenordnung eines Grades in 12 Stunden, also nur den 4. Teil des beobachteten Wertes.

Im oberen Teil der Masse ist es nicht möglich etwa vorhandene Wärmequellen zu erkennen. Es fällt nur auf, daß sich der Schirm der polaren Tropopause zwischen Zeit 1 und Zeit 2 nördlich und westlich der Masse auflöst, während im Süden noch bis Zeit 4 Reste von ihm zu erkennen sind. Das kann, da die Tropopause ein relatives Temperaturminimum ist, eine Erwärmung bedeuten. Es kann aber auch auf verschieden starker vertikaler Streckung der darüber und darunter liegenden Massen beruhen. Im Innern von A verschwindet die Fläche b hauptsächlich infolge der Abkühlung an der Fläche a.

Was zwischen Zeit 3 und 4 im Temperaturfeld der Masse A erfolgt, kann man als Alterung deuten. An der Untergrenze setzt wieder eine leichte Erwärmung ein. Dadurch wird die Fläche a ziemlich verwischt, es entsteht eine tiefer liegende, aber auch nicht viel deutlichere Grenzfläche. Es ist vielleicht bemerkenswert, daß das Temperaturfeld der Höhenzyklone zu *der* Zeit am schärfsten ausgeprägt ist, da die Zyklone gerade erst als selbständiges Gebilde zu existieren beginnt.

Zusätzliche Kräfte im Bereich der Masse A.

Zur Zeit 1 lag die Masse A mitten in der NW-Strömung, besaß also vermutlich im Mittel nur eine geringe relative Rotation. Zur Zeit 4 betrug, wie schon erwähnt, die Absolutrotation im Kern $27 \cdot 10^{-5}$ sec^{-1}. Ihre Zunahme auf etwa das Dreifache ist ohne die Mitwirkung eines zusätzlichen Kraftfeldes nur möglich, wenn sich die Masse in vertikaler Richtung auf das dreifache Druckintervall streckt. Das ist sicher nicht der Fall. Die Untergrenze ist die Fläche a, als Obergrenze kann man, ohne sehr fehl zu gehen, ständig dieselbe isentrope Fläche, z. B. $\vartheta = 60°$, wählen. Denn bei der hohen Stabilität der Stratosphäre verschiebt auch eine Erwärmung um mehrere Grad die isentropen Flächen nicht sehr im Massenfeld. Aus den Vertikalschnitten entnimmt man, daß Unter- und Obergrenze ständig etwa im selben Abstand liegen. Es müssen also zusätzliche Kräfte vorhanden sein.

Exakt dargestellt wird dieser Zusammenhang durch die vorticity-Gleichung. Wir benützen sie in der Form

$$\frac{d}{dt}(f + \text{Rot } \mathbf{v}) = -(f + \text{Rot } \mathbf{v}) \text{ Div } \mathbf{v} + \text{Rot } \mathbf{E}.$$

Rot ist die vertikale Komponente der Rotation, Div die horizontale Divergenz, **E** die Summe der eingeprägten Kräfte.

Die Gleichung gilt auch, wenn man die Operatoren Rot und Div auf die Felder einer Isentropen-Karte anwendet. Ja, sie gilt hier sogar eher als bei der normalen Anwendung, wo Rot und Div aus den Werten der Horizontalen bestimmt werden. Denn hierbei treten, wenn die horizontale flüssige Fläche nicht ständig horizontal bleibt, zusätzliche Glieder in der vorticity-Gleichung auf, die bei starken vertikalen Windgradienten erhebliche Beträge erreichen können. Von diesen »Kippgliedern« bleibt man frei, wenn man Winde und Kräfte in den Punkten einer massenfesten (flüssigen) Fläche vertikal projiziert und die vorticity-Gleichung auf die Projektionen anwendet. Denn für die projizierten Felder gelten dieselben Gesetze wie für die 2-dimensionale Strömung in einer Ebene, und in einer ebenen Strömung gilt obige Gleichung streng. Die Isentropenkarten sind also, vorausgesetzt, daß die isentropen Flächen massenfest sind, besonders geeignet für die Anwendung der vorticity-Gleichung.

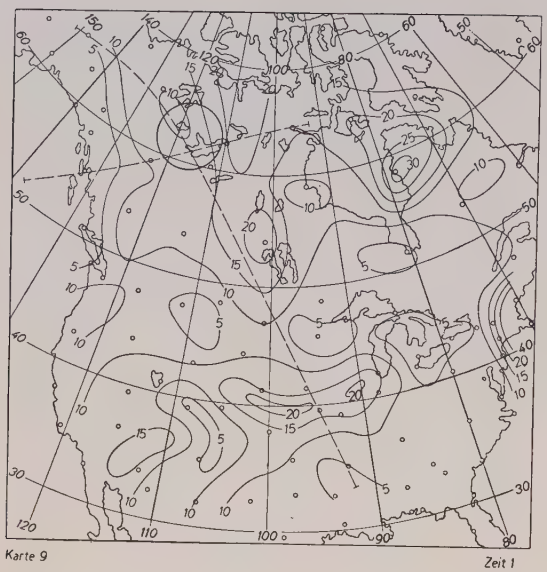
Ein weiterer Vorteil ist dabei, daß in dem Glied Rot **E** der Anteil der Druckgradientkraft verschwindet. Denn ihre horizontale Komponente hat bei der Darstellung in Isentropen-Karten ein Potential in Gestalt der Größe Ψ. Ein von Null verschiedener Wert von Rot **E** kommt also ganz auf das Konto von zusätzlichen Kräften.

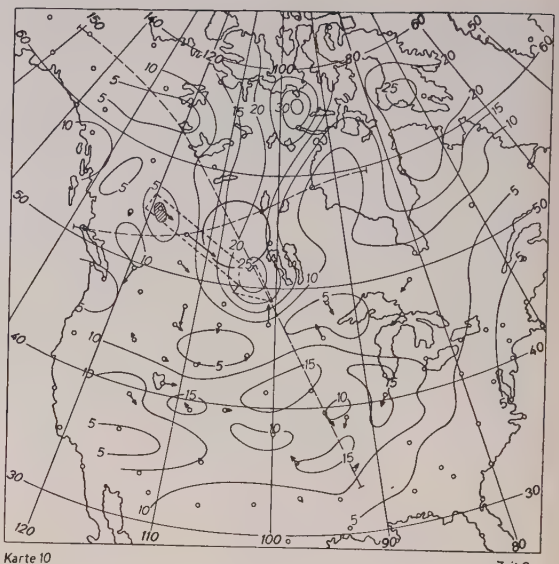
Wie immer, muß man zur numerischen Rechnung auf den geostrophischen Wind zurückgreifen. Aus den Isentropen-Karten läßt sich Rot $\mathbf{v}_g$ (isentrop) auf ähnliche Weise gewinnen wie Rot $\mathbf{v}_g$ (horizontal) aus den Isobar-Karten. Man hat nur in der bekannten Gleichung Φ durch Ψ zu ersetzen, d. h. es ist

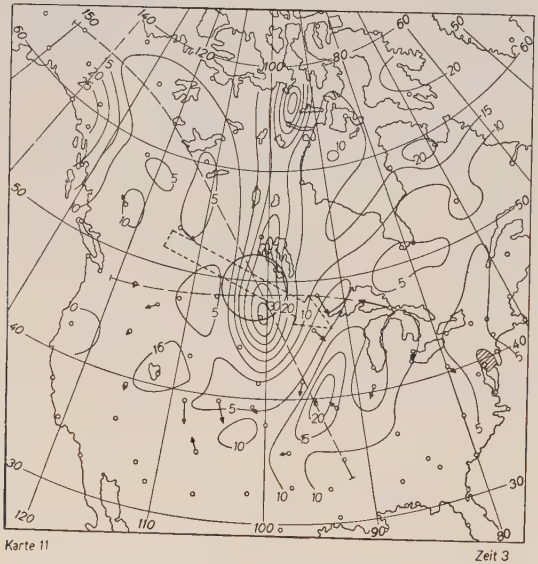
$$\text{Rot } \mathbf{v}_g = \frac{\triangle \Psi}{f}.$$

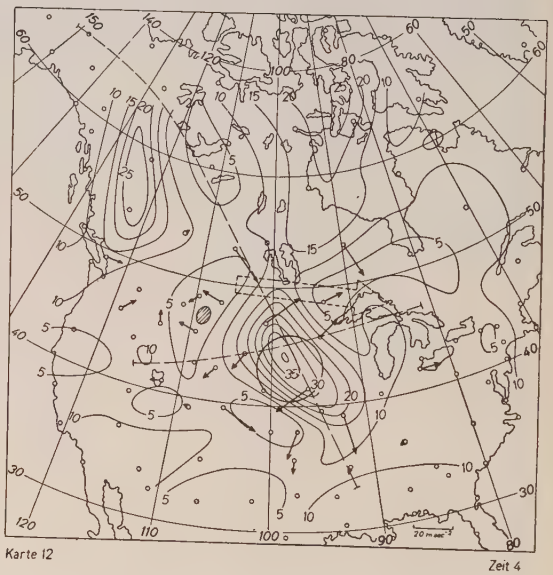
Infolgedessen konnte die Berechnung nach demselben Verfahren erfolgen, das Bolin und Charney (1951) bei Isobar-Karten angewandt haben. Bezüglich der Einzelheiten des Verfahrens sei auf diese Arbeit verwiesen.

Das Ergebnis ist in den Karten 9—12 darge-









Karten 9—12: Isentropenkarten der geostrophischen Absolutrotation für die 35°-Fläche, Angaben in 10^{-5} sec^{-1}. Gestrichelt: Spurlinien der Vertikalschnitte. Schraffiert: Gebiete mit negativer Absolutrotation. Über die Pfeile siehe Text.

stellt. Sie geben die Absolutrotation $f+\mathrm{Rot}\,\mathbf{v}_g$ auf der 35°-Fläche.

Die geostrophische Absolutrotation weicht oft erheblich von der des wahren Windes ab. Bei starker Krümmung, besonders im Innern von Zyklonen, kommt man dem wahren Wert sicher näher, wenn man statt des geostrophischen den Gradientwind benützt. Zwischen der Absolutrotation des Gradientwinds $\mathbf{v}_i$ und der des geostrophischen Windes $\mathbf{v}_g$ besteht, wie im Anhang bewiesen wird, folgende Beziehung

$$f+\mathrm{Rot}\ \mathbf{v}_i=\frac{2fv_g+rf\,(f+\mathrm{Rot}\ \mathbf{v}_g)}{\pm\sqrt{r^2f^2+4rfv_g}}.$$

In dieser Formel bedeutet r den Krümmungsradius der Ψ-Linie, bei zyklonaler Krümmung positiv, bei antizyklonaler negativ gerechnet. Das Vorzeichen der Wurzel ist jeweils dasselbe wie das von r.

Der Unterschied zwischen den beiden Absolutrotationen ist am größten im Zentrum einer Zyklone. Setzt man hier

$$f+\mathrm{Rot}\ \mathbf{v}_g=n\cdot f,$$

wo n eine Zahl >1, so ist

$$f+\mathrm{Rot}\ \mathbf{v}_i=\sqrt{2n-1}\cdot f.$$

Das bedeutet z. B., daß dem Maximum der geostrophischen Absolutrotation in Karte 12, das etwas über $40\cdot10^{-5}$ sec^{-1} beträgt, eine Absolutrotation des Gradientwinds von nur $27\cdot10^{-5}$ sec^{-1} entspricht, ein Wert, der zufällig genau mit dem oben angegebenen, aus den Höhenwinden berechneten Wert übereinstimmt.

Reduziert man auch die Angaben der anderen Karten nach obiger Gleichung, so werden die Maxima in den Karten 10 und 11 (28 und $37\cdot10^{-5}$) ebenfalls stark abgeschwächt und außerdem gegen das Zentrum der Kreise verschoben. Bei Karte 9 ändert sich dagegen in der Nähe des Kreises fast nichts. Für die 4 Zeiten erhält man als ungefähre Absolutrotation im Kern der Masse A die in der ersten Zeile der Tabelle stehenden Werte.

Tabelle

Zeit	1	2	3	4	
$f+\mathrm{Rot}\,\mathbf{v}$	10	17	28	27	10^{-5} sec^{-1}
D	190	190	160	200 bis 250	mb
Rot **E**	+ 16	+ 35	− 16 bis − 30		10^{-10} sec^{-2}

Sehr genau sind die Werte nicht, schon deswegen, weil ja auch der Gradientwind vom wahren Wind noch abzuweichen pflegt.

Div **v** berechnen wir aus der Änderung der Druckdifferenz D zwischen Ober- und Untergrenze von A. Es ist näherungsweise[1]

$$\mathrm{Div}\ \mathbf{v}=-\frac{1}{D}\frac{dD}{dt}.$$

Als Obergrenze wählen wir, wie gesagt, die 60°-Fläche, als Untergrenze die Fläche a. Gemessen wird D im Mittelpunkt der Kreise. Die gefundenen Werte stehen in der zweiten Zeile der Tabelle. Zur Zeit 4 ist die Lage der flüssigen Fläche a zweifelhaft. Deswegen sind 2 Grenzen angegeben, zwischen denen der wahre Wert vermutlich liegt.

Setzt man Div **v** in die vorticity-Gleichung ein, bildet das Zeitintegral und ersetzt die Integranden durch die jeweiligen Mittelwerte, so erhält man für das mittlere Rot **E** in den 3 Zeitabschnitten die in der 3. Zeile angegebenen Werte. Die beiden Zahlen im letzten Zeitabschnitt entsprechen den beiden Werten von D zur Zeit 4.

Wenn die angegebenen Zahlen richtig sind, so hat sich folgendes abgespielt: Zwischen Zeit 1 und Zeit 3 hat die Masse A durch ein zusätzliches Kraftfeld ein zyklonales Drehmoment erfahren. Ein solches Drehmoment läßt an sich eine Luftmasse horizontal divergieren. Im ersten Zeitabschnitt tritt aber keine Divergenz ein, offenbar weil die Masse zunehmend von subtropischer Luft umschlossen wird. Die Abnahme in der mittleren Stabilität der Umgebung verhindert eine Divergenz. Erst im zweiten Zeitabschnitt, wo das Drehmoment verdoppelt ist, divergiert die Luftmasse etwas.

Mit Zeit 3 wird das Drehmoment plötzlich antizyklonal. Drehmoment und neue Umgebung wirken nun im gleichen Sinn. Die Masse konvergiert. Zufällig bleibt die Absolutrotation dabei annähernd erhalten.

Daß das Drehmoment gerade in dem Zeitpunkt, wo die Zyklone sich eben zu bilden beginnt, das Vorzeichen wechseln soll, mutet zunächst ziemlich unwahrscheinlich an. Es erfährt aber im folgenden eine gewisse Bestätigung.

Wenn von zusätzlichen Kräften derart starke

[1] Exakt gilt die Gleichung nur für isentrope Schichten infinitesimaler Dicke ($D\to0$).

Drehmomente ausgeübt werden, so sollte man meinen, daß sich die Kräfte selbst durch Abweichungen des Windes vom geostrophischen Wind bemerkbar machen. Von diesem Gedanken ausgehend, wurde überall dort, wo Windmessungen vorlagen, diejenige Kraft bestimmt, die den horizontalen Komponenten von Corioliskraft und Druckkraft die Waage hält. Unter der Voraussetzung, daß die Luftbewegung unbeschleunigt ist, ist diese Kraft gleich der Summe der zusätzlichen Kräfte. Der einzelne, auf diese Weise gewonnene Kraftvektor ist im allgemeinen erheblich falsch, da die gesamte Beschleunigung des Luftteilchens oft von derselben Größenordnung ist wie die aus den Zusatzkräften herrührende. Wendet man aber das Verfahren auf alle Windmessungen eines großen Gebietes an, so kann man hoffen, großräumige Zusatzkraftfelder an einer Vorzugsrichtung der Vektoren zu erkennen.

In den Karten der Absolutrotation (9—12) sind als Pfeile die auf diese Weise gewonnenen »Zusatzkräfte« eingetragen. Rechts unten auf Karte 12 ist der Maßstab für ihre Stärke angegeben[1]. Der erste Eindruck ist der, daß die Pfeile ganz unregelmäßig durcheinander laufen. Die meisten sind so klein, daß sie wohl im Wesentlichen auf der vernachlässigten Beschleunigung oder auf Fehlern in Grad Ψ beruhen. Eine Vorzugsrichtung bei größeren Beträgen kann man nur in den Karten 10 und 12 feststellen, und zwar längs der großen gestrichelten Pfeile. Zur Zeit 1 (Karte 9) liegen keine Windmessungen im Umkreis der Masse A. Auch zur Zeit 3 fehlen die Messungen in ihrer unmittelbaren Nähe, doch sprechen drei Pfeile in Karte 11 dafür, daß das Kraftfeld auch hier vorhanden ist und gerade etwa durch die Mitte der Masse A führt.

Größere Pfeile von stark wechselnder Richtung liegen noch zur Zeit 4 in der Südhälfte der Zyklone. Vermutlich beruhen sie nicht so sehr auf zusätzlichen Kräften, als auf besonders hohen Beschleunigungen. Denn wie man aus den Windpfeilen in Karte 4 sieht, läuft hier die Luft nicht gleichmäßig um den Zyklonenkern, sondern in ziem ich eckigen Bahnen.

Den großen gestrichelten Pfeilen würde man vielleicht keine Realität zubilligen, wenn sie nicht mit unserem obigen Ergebnis im Einklang wären. Das Vorzeichen der in der Tabelle stehenden Werte von Rot **E** entspricht genau der jeweiligen Lage des Kraftfeldes. Auch die Größenordnung stimmt. Der Pfeil an der Station Edmonton in Karte 2 – es ist die Station, deren Windmessung erst vermuten ließ, die Masse A käme mehr aus Norden – entspricht einer Beschleunigung von $19 \cdot 10^{-4}$ m sec^{-2}. Nimmt man an, daß das Kraftfeld linear nach NE abnimmt und auf der Gegenseite von A den Wert 0 erreicht, so folgt daraus ein Rot **E** von $24 \cdot 10^{-10}$ sec^{-2}, ein Wert, der gut mit den in der Tabelle stehenden übereinstimmt.

Noch ein Drittes spricht für die Existenz des Kraftfeldes. Das ist die Bahn der Masse. Wir hatten zwar gefunden, daß die Bahn dem geostrophischen Wind folgte. Das bedeutet, da die Bahnbeschleunigung nur gering ist, daß im Zentrum der Masse keinerlei zusätzliche Kräfte wirksam sind, im Gegensatz zu dem gestrichelten Pfeil in Karte 11. Man sieht aber, besonders aus den Karten 2 und 3, daß die Bestimmung der geostrophischen Bahn überaus unsicher ist, und daß ihre Übereinstimmung mit der wirklichen Bahn kaum den Schluß rechtfertigt, die Masse sei geostrophisch gezogen. Auf der andern Seite gewinnt man aus den Schnitten den deutlichen Eindruck, daß die Masse aus der allgemeinen Strömung nach rechts ausschert. Das ist aber mit unserem Kraftfeld gut zu vereinen. Denn in einem zusätzlichen Kraftfeld werden die Massen nur kurze Zeit in Richtung des Feldes zusätzlich beschleunigt. Dann werden sie durch die Corioliskraft nach rechts abgelenkt und erhalten eine zusätzliche Geschwindigkeit quer zum Kraftfeld. Auf diese Weise wird eine Masse, die ursprünglich links vom Kraftfeld lag, auf die rechte Seite hinübergeschafft. Gerade das zeigen die Karten.

Um die vertikale Erstreckung des zusätzlichen Kraftfeldes kennenzulernen, wurden die »Zusatzkräfte« auch für die Höhenwinde der 700-, 500-, 400- und 300 mb-Fläche berechnet. Es ergab sich, daß das Kraftfeld zur Zeit 2 und 3 bis ins 700 mb-Niveau hinunterreichte, allerdings dort kaum noch zu erkennen war. In den Schnitten 6 und 7 sind diejenigen Winde schwarz ausgefüllt, an denen sich das Kraftfeld bemerkbar machte. Zur Zeit 4 war es noch bei 500 mb vorhanden, bei 700 mb dagegen nicht mehr. Nach oben war eine Begrenzung aus Mangel an Windmessungen nicht festzustellen.

[1] Leider ist dabei durch ein Versehen der Faktor 10^{-4} vergessen worden.

Schluß

Der Nachweis starker zusätzlicher Kräfte bei kurzfristigen Vorgängen ist für unsere Vorstellung von der Dynamik der Atmosphäre so bedeutsam, daß man jedem derartigen Ergebnis mit großer Skepsis entegentreten wird. Auch in unserem Fall wird man sich fragen, wie weit das Ergebnis reell und ob es nicht teilweise durch falsche Analyse vorgetäuscht ist. Leider läßt sich dazu nicht mehr sagen, als was schon gesagt wurde. Natürlich gibt es erhebliche Fehlerquellen. Die Bestimmung von Rot **v** ist sehr unsicher. Vielleicht verläuft auch die Bahn der Masse etwas anders. Eines erscheint aber sicher: ganz ohne zusätzliche Kräfte geht es nicht. Wenn man Konstanz des potentiellen Wirbelwerts verlangt, hätte die Masse anfangs weiter östlich liegen und ihre Bahn um 10 bis 20° vom geostrophischen Wind abweichen müssen. Das wäre aber wiederum nur unter Mitwirkung erheblicher zusätzlicher Kräfte möglich gewesen.

Weitere Sicherheit kann man nur auf *einem* Wege gewinnen. Man muß weitere Fälle analysieren. Mit einer größeren Anzahl Analysen könnte man wohl entscheiden, was an der Einzelanalyse richtig und was fraglich ist. Man würde erkennen, welche Erscheinungen für das Cut-off wesentlich sind und welche zufällig. Man denke hier z. B. an die Abkühlung an der Untergrenze der Masse. Man würde schließlich auf diese Weise ein wesentlich vollständigeres Bild des ganzen Vorgangs erhalten.

Zum Schluß erhebt sich natürlich noch die Frage nach der *Natur* des zusätzlichen Kraftfeldes. Wenn zusätzliche Kräfte in der Atmosphäre nachgewiesen werden, so pflegt man die Reibung und den Austausch dafür verantwortlich zu machen. Leider sind wir noch weit von einer gesicherten Theorie des Austauschs entfernt. Ähnliche Vorgänge wie die hier geschilderten haben zur Vorstellung des Gleitaustausches geführt, der in den horizontalen oder genauer in den isentropen Flächen vor sich geht. Wählt man als Austauschgröße mit Rossby (1947) die Absolutrotation oder mit Raethjen (1949) das absolute Rotationsmoment, so treten, sowie der Gleitaustausch in einem Gebiet einsetzt, an den Rändern des Gebiets zusätzliche Beschleunigungen auf.

Ob das zusätzliche Kraftfeld in unserem Fall auf derartigen Vorgängen beruht, läßt sich bei der Spärlichkeit der aerologischen Meldungen nicht entscheiden. Wenn der Austausch in irgend einer Form die Ursache ist, muß an anderer Stelle ein Kraftfeld entgegengesetzter Richtung auftreten. Denn die durch den Austausch bewirkte Impulsänderung muß im Mittel über den ganzen Austauschbereich verschwinden. Ein solches Kraftfeld ist in unserem Fall nicht nachzuweisen. Es kann natürlich sein, daß der Impuls, den die Luft in unserem Kraftfeld erfährt, derart ausgedehnten Massen entzogen wird, daß der Entzug im Windfeld nicht bemerkt wird. Vielleicht kann man auch auf diese Fragen aus einer größeren Anzahl von Analysen eine Antwort erhalten.

ANHANG

Der Zusammenhang zwischen der Absolutrotation des geostrophischen Windes und der des Gradientwindes

Zwischen den beiden Windstärken besteht die Beziehung

$$\frac{v_i^2}{r} + f v_i = f v_g , \tag{1}$$

wo r bei zyklonaler Krümmung positiv, bei antizyklonaler Krümmung negativ zu rechnen ist. Die Absolutrotation des Gradientwindes ist

$$f + \mathrm{Rot}\ \mathbf{v}_i = f + \frac{v_i}{r} + \frac{\partial v_i}{\partial r}, \tag{2}$$

die des geostrophischen Windes

$$f + \mathrm{Rot}\ \mathbf{v}_g = f + \frac{v_g}{r} + \frac{\partial v_g}{\partial r}. \tag{3}$$

Aus (1) folgt durch Differentiation nach r

$$\frac{\partial v_i}{\partial r} = -\frac{f v_i - f v_g - r f \dfrac{\partial v_g}{\partial r}}{2 v_i + r f}. \tag{4}$$

Setzt man dies in (2) ein, so ergibt sich mit Hilfe von (1)

$$\left.\begin{aligned} f + \mathrm{Rot}\ \mathbf{v}_i &= \frac{3 f v_g + r f^2 + r f \dfrac{\partial v_g}{\partial r}}{2 v_i + r f} = \\ &= \frac{2 f v_g + r f (f + \mathrm{Rot}\ \mathbf{v}_g)}{2 v_i + r f} \end{aligned}\right\} \tag{5}$$

Um den Nenner zu bestimmen, hat man Gleichung (1) nach v_i aufzulösen. Der Gradientwind entspricht der Lösung

$$2v_i = -rf \pm \sqrt{r^2f^2 + 4rfv_g}\,, \qquad (6)$$

wo vor der Wurzel das positive Vorzeichen bei zyklonaler, das negative bei antizyklonaler Krümmung gilt. Folglich ist

$$f + \mathrm{Rot}\ \mathbf{v}_i = \frac{2fv_g + rf\,(f + \mathrm{Rot}\ \mathbf{v}_g)}{\pm\sqrt{r^2f^2 + 4rfv_g}}\,. \qquad (7)$$

Das ist der gesuchte Zusammenhang.

LITERATUR

Bolin, B. and Charney, J. G. (1951): Numerical tendency computations from the barotropic vorticity equation. *Tellus* **3**, 248—257.

Elsasser, W. M. (1942): Heat transfer by infrared radiation in the atmosphere. *Harvard Meteorological Studies*, Nr. 6.

Ertel, H. (1942): Ein neuer hydrodynamischer Wirbelsatz. *Meteor. Z.*, **59**, 277—281.

Kleinschmidt, E. (1950): Über Aufbau und Entstehung von Zyklonen, 1. und 2. Teil. *Meteor. Rundsch.* **3**, 1—6 und 54—61.

Montgomery, R. B. (1937): A suggested method of representing gradient flow in isentropic surfaces. *Bull. Am. Met. Soc*, **18**, 210—212.

Raethjen, P. (1949): Zyklogenetische Probleme. *Archiv f. Meteor., Geophysik u. Bioklimatologie*, Serie A, Band I, 295—346.

Rossby, C.-G. (1947): On the distribution of angular velocity in gaseous envelopes . . ., *Bull. Am. Met. Soc.*, **28**, 53—68.

Upper Mean Flow over the North Atlantic during January 1952

By WILLIAM HUBERT and YNGVE DAGEL, University of Stockholm

(Manuscript received 30 July 1954)

Abstract

Three mean atmospheric cross sections, one along the eastern coast of North America near longitude 80° West, one extending from Hudson's Bay out over the western Atlantic and the third along the western coasts of Europe and North Africa for January 1952 are presented. Four distinct currents are discussed in some detail; they are:

A. A polar stratospheric jet stream which is most clearly defined over Scandinavia where it shows up as a separate current centered near latitude 62° North. The maximum winds are above the 100-mb level.

B. A tropospheric jet stream in middle latitudes which weakens appreciably between the eastern coast of North America and western Europe but can be traced through all three cross sections (central potential temperature about 335° A).

C. A well-defined maximum over North Africa at latitude 24° North which has no clear connection with any maxima over North America and may form over the Atlantic Ocean. This current is centered at a potential temperature of approximately 345° A.

D. A mean low-latitude west-wind maximum over Panama at 45,000 feet near latitude 11° North. Mean observed wind profiles at Albrook Field, Panama and at Swan Island indicate that this really is a separate current and not the south edge of the middle-latitude jet stream. The central potential temperature in this maximum is about 355° A.

Introduction

A number of mean atmospheric cross sections through North America, in particular along longitude 80° West, have now been published. One of the primary reasons why this longitude has received so much attention is the excellent network of aerological stations which is found here. However, the network over western Europe is even more dense in places, and yet there has been a hesitancy on the part of most synoptic meteorologists to construct monthly-mean cross sections along the eastern side of the Atlantic.

Over the ocean itself, the limiting factor in cross-section analysis has obviously been the scarcity of reporting stations. There are, nevertheless, enough weather ships to allow construction of mean sections over part of the North Atlantic. Since we have no upper-air stations at low latitudes in the middle of the ocean, it is only possible to draw conclusions about the upper flow here through the study of sections upstream and downstream.

The work described in this paper was undertaken in an attempt to determine whether or not large scale features in the general circulation could be traced from one side of the Atlantic to the other. The jet stream maps of NAMIAS and CLAPP (1951) indicate a lack of continuity in this region; the authors point out, however, that large changes may be necessary as more data become available.

Mean Flow at 200 MB

In figure 1 is presented a chart of the mean 200-mb contours over the Atlantic Ocean and adjacent land areas during January 1952. The analysis is based upon mean values of the 0300 GCT radiosonde reports from those stations indicated by black dots. The larger dots give the locations of the stations used in the cross sections which will be discussed later.

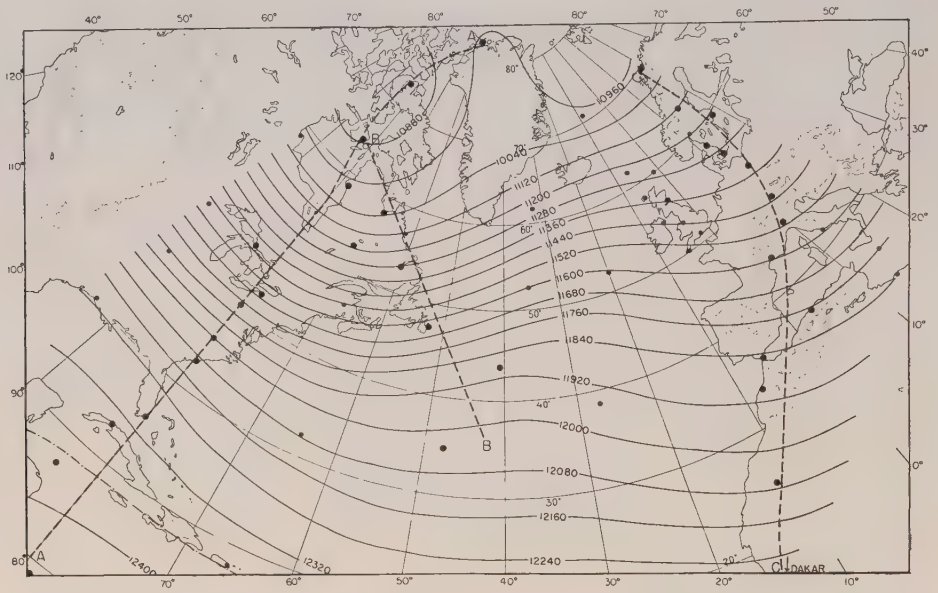


Fig. 1. Mean 200-mb chart at 0300 GCT for January 1952. Black dots represent stations used in analysis. Thin solid lines, 200-mb contours at intervals of 80 meters. Heavy dashed lines, locations where mean cross sections were constructed.

During this month the major trough off the western coast of the United States was well developed in the mean thereby giving more of a westerly flow across the central part of the country than is usual. This in turn must be connected in some way with an inhibition of deep polar outbreaks east of the Rocky Mountains and therefore relatively high index flow.

Perhaps the most predominant feature of the mean flow pattern over the ocean is the semi-permanent long wave which has a wave length roughly equal to the width of the Atlantic. The wave is fairly symmetrical; the troughs are slightly inland from the eastern and western shores, and the ridge is located approximately in mid-ocean. This is, of course, no coincidence, for the role of the earth's orography in determining the location of the semi-permanent centers of action is fairly well known (eg. Bolin 1950).

From a casual inspection of figure 1, one can also see that the broad current maximum (jet stream) over the United States must weaken considerably downstream. This is evident from the marked diffluence of the 200-mb contours, especially over western Europe. On the eastern side of the North Atlantic the contour gradient is quite weak in the vicinity of the Azores; farther south, however, there is obviously a second current maximum. The confluence of the contours between latitudes 20° and 30° North along the west coast of Africa indicates that this stream has a maximum intensity farther downstream (see Bannon 1954).

The average jet stream charts by Namias and Clapp and the mean January isotachs at 500 mb by Mintz and Dean (1951) both indicate that the stream passing over North Africa has no direct connection with that over the eastcentral United States. It is furthermore interesting to note from the isotach analyses referred to above that one jet dies

out while the other intensifies at the longitude where the occurrence of blocking action is a maximum (Rex 1950).

A "climatological jet" which results from averaging over a period as long as a month is frequently unrepresentative of day-to-tay conditions. We know, for instance, that double jet streams over the United States are not exceptional (see eg. Cressman 1950); it may be that the southernmost of these occasionally continues eastward to the south of Bermuda and the Azores. If such is the case, the mean polar-front jet stream and the mean subtropical jet apparently merge over the Eastern United States to form one broad maximum, and it is only farther downstream that separate current maxima again become distinguishable. The reverse of this may be seen over eastern Asia in winter; one mean jet over south China and another north of the Himalayas (Yeh 1950) merge, in the vicinity of Japan, into a single maximum of extremely high winds (Mohri 1953).

Mean Cross Section for Eastern North America

A mean cross section for January 1952 from Alert on northern Greenland along the dashed line A—A (figure 1) to Albrook Field, Panama Canal Zone is shown in figure 2. The mean heights and temperatures were determined at 1000, 850, 700, 500, 400, 300, 200, and 100 mb from the 0300 GCT upper-air soundings. Nighttime observations were used throughout this study in order to reduce radiational effects. All soundings which were not complete up to 100 mb were built up to the highest standard level reached by more than half of the soundings for the month. This was done by carefully estimating the missing portion of the p, T curve using analysed charts based upon nearby stations as well as 1500 GCT observations from the same station. Geostrophic winds normal to the section were computed from height profiles for each constant pressure surface.

We see a single broad jet stream (B) over

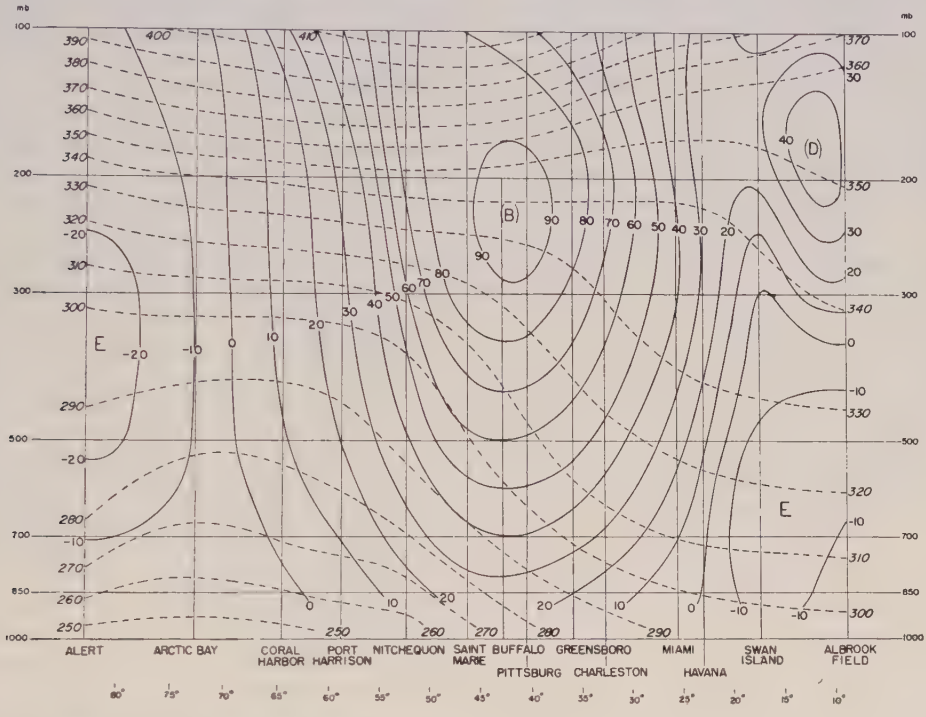


Fig. 2. Mean atmospheric cross section at 0300 GCT for January 1952 from Alert (Greenland) to Albrook Field (Panama) along line A-A in figure 1. Solid lines, geostrophic wind speed (knots) normal to section. Dashed lines, potential temperature (Deg. A.).

the eastern United States centered at latitude 41° North with its core near 225 mb at a potential temperature of approximately 335° A. The maximum geostrophic wind (95 knots) agrees well with the winter cross section of Hess (1948) along this longitude; however, the latitude of the center is about 5° farther north in this case, as is characteristic of high-index flow. North of 70° latitude the flow

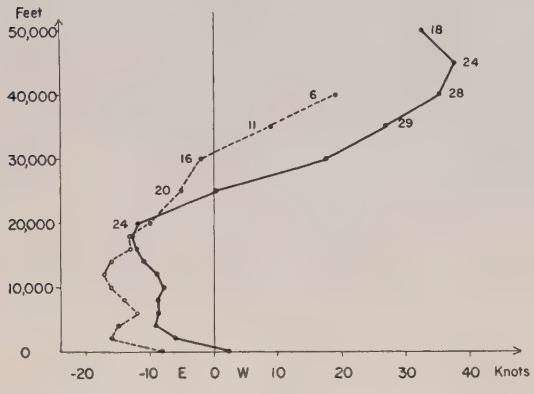


Fig. 3. Mean observed zonal wind profiles (knots) at Albrook Field, Panama (solid line) and Swan Island (dashed line) for January 1952.

was easterly at all levels up to 100 mb although the easterlies decreased in intensity above the polar tropopause.

The most unusual feature in this cross section compared to other mean sections along the same longitude is the presence of a narrow westwind maximum (D) centered near 11° North. Geostrophic wind computations at 200 mb yielded a maximum wind of 47 knots during this month whereas Hess found a mean for four winters (Jan.—Feb., 1942—45) of only 10 knots at the same altitude. The reality of this low-latitude westerly current is confirmed by the wind observations from Albrook Field at latitude 9° North and Swan Island at latitude 17° North. In figure 3 are shown the mean observed zonal wind profiles at Albrook and Swan Island (solid and dashed lines respectively) for the month of January 1952. Wind data at 0300 GCT were used for the most part; however, it was necessary to add some 1500 GCT observations in order to have one ascent for each day of the month.

Although the number of observations from high levels at Swan Island leaves much to be desired, it sems clear that the westerlies were weaker there than over Albrook (in agreement with geostrophic computations). The origin of these strong winds at the latitude of Albrook is difficult to explain. Assuming conservation of angular momentum, a ring of air at rest at the equator would have a relative speed of about 25 knots if displaced to latitude 9° North. This value is considerably less than the observed maximum of 38 knots at Albrook, so it is obvious the observed winds cannot be explained by conservation of momentum. Either the current observed here is the result of some relatively local acceleration, in which case it is not a hemispheric phenomenon, or it is part of a ring extending all the way around the earth, in which case a method whereby momentum can be transferred normal to the flow is required. A mechanism wherein the latter is possible has not as yet been developed.

Some ascents which reached very high levels at Albrook indicate that the flow is again easterly in the stratosphere with a maximum around 80,000 feet. Riehl (1954) has classified this type of wind profile, at least that part in the troposphere, as E_d W meaning deep easterlies with strong westerlies at high levels. It may be that the low-latitude stream found here is well developed only during high-index conditions (when the middle-latitude jet is farther north than usual).

Mean Cross Section for Northwest Atlantic

In figure 4 is presented a mean cross section from Coral Harbor, Canada to ocean weather station "E" along the heavy dashed line B—B in figure 1. The broad jet stream which was centered near 41° North over eastern North America crosses longitude 50° West at 47° North latitude; the maximum geostrophic wind in the center has decreased about 10 knots between the two sections. The core is located at the same pressure and has the same potential temperature as farther upstream (225 mb and 335° A). The strongest horizontal wind shear is found on the south side of the jet which probably indicates that the latitudinal fluctuations of the jet from one day to the next are blocked to the south by the relatively stationary Bermuda high. The vertical wind shear in this section is more concentrated in

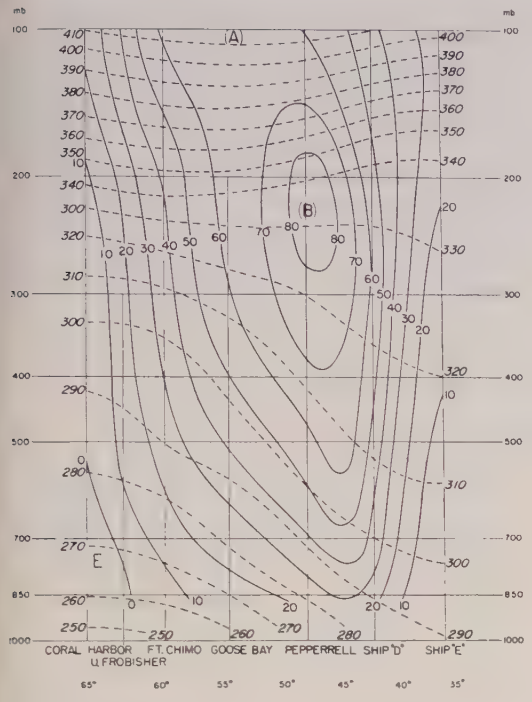


Fig. 4. Mean atmospheric cross section at 0300 GCT for January 1952 from Coral Harbor (Canada) to Ocean Weather Ship "E" along line B-B in figure 1. Legend same as in figure 2.

low levels whereas at longitude 80° West it was fairly uniform throughout the troposphere.

Mean Cross Section for Western Europe and Africa

Figure 5 is a mean cross section for January 1952 from Tromsö, Norway to Dakar, French West Africa. The section was constructed along the curved, dashed line C—C in figure 1. For Gibraltar it was necessary to utilize soundings made at 1500 GCT. At Fort Trinquet soundings are made only every second day; however, this inhomogeniety is probably not serious because of the relatively small changes which occur here.

In this section the middle-latitude jet stream, which was the most prominent feature upstream, has weakened much more and appears over the Mediterranean as a wide maximum with a wind speed of only 54 knots in the core. The jet center is again at approximately the same pressure and potential temperature so that it appears reasonable to consider this stream as being continuous across the Atlantic during this particular period.

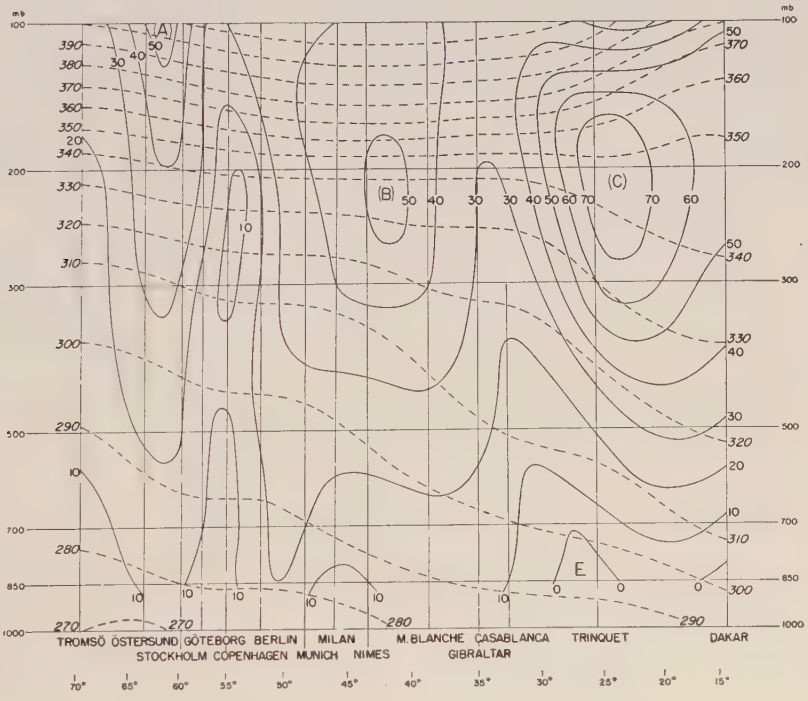


Fig. 5. Mean atmospheric cross section at 0300 GCT for January 1952 from Tromsö (Norway) to Dakar (French West Africa) along line C-C in figure 1. Legend same as in figure 2.

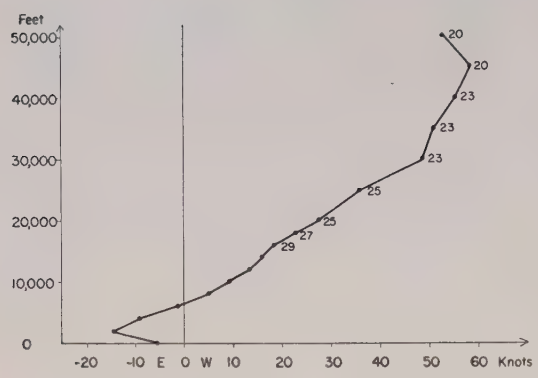


Fig. 6. Mean observed zonal wind profile (knots) at Dakar, French West Africa, for January 1952.

The unusually light winds in the vicinity of latitude 55° North are the result of cut-off lows which persisted over southern Europe during a good part of the month. At Berlin the observed winds at 300 mb had an easterly component from the 19th through the 24th of January. Such persistent anomalies are a common occurrence over Europe, and the monthly mean pattern shown in figure 5 undoubtedly differs from the long-term mean for winter conditions.

A wind maximum in the polar stratosphere near latitude 62° North (A) is visible in this cross section. Mean geostrophic winds of 56 knots were computed at 100 mb, and the thermal field indicates that the winds continue to increase above this level. This polar stratospheric jet stream is found only in mid-winter near the boundary of polar night; the latitudinal variation in direct solar radiation must be very important at high levels during this time of year (PALMÉN 1934).

The strongest winds in this cross section (77 knots) are over North Africa near 24° North. Both the location and intensity of this jet agree favourably with the January means of Namias and Clapp. The central potential temperature of this stream is about 345° A, and the core is located near 210 mb. JAMES (1951), in an earlier mean cross section for February 1951 along the Greenwich meridian, found that the west winds increased with height over North Africa, but he obtained no closed maximum.

Geostrophic computations at both 200 and 100 mb gave some indication that the winds at these levels increase again south of latitude 16° North. Perhaps the west-wind maximum which was found over Panama continues across the Atlantic at approximately the same latitude. The mean observed zonal winds at Dakar, figure 6, substantiate this possibility in so far as the maximum westerlies over Albrook Field and over Dakar occur at the same height (45,000 feet). The time-section from the Finnish Atlantic expedition in October—November, 1939 also shows a westerly current at almost the same altitude and latitude (see VUORELA, 1948).

Conclusions

An attempt has been made in figure 7 to summarize the results of the three mean cross sections by showing the probable locations of the four currents described above. It is possible that the jet stream over North Africa is continuous across the Atlantic and merges with the middle-latitude jet over the east coast of North America; however, since this has not been definitely established, it is depicted as forming over the ocean. The temporary addition of two weather ships at low latitudes in the North Atlantic could settle this question. This could conceivably be accomplished during the next International Geophysical Year.

More investigation of the low-latitude stream over Panama seems warranted. It would be interesting to know whether a maximum here is a relatively rare occurrence and whether the phenomenon is local or hemispheric in scale.

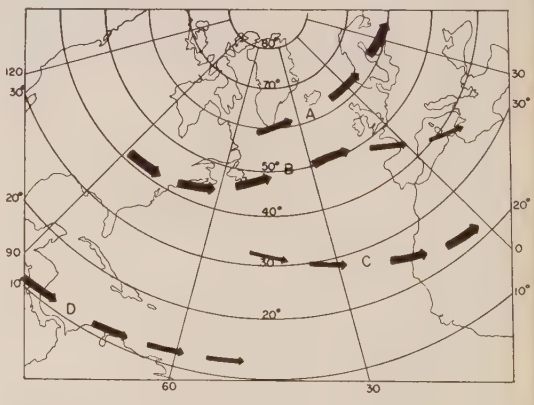


Fig. 7. Attempt to show probable location of the four jet streams found in the mean cross sections for January 1952.

The work involved in constructing these cross sections has emphasized three important points in connection with the collection and handling of data, (1) the need for greater uniformity of radiosonde types, especially in Europe, (2) the need for a single radiosonde code instead of the present feet—°F in one country and meters—°C in another and (3) the desirability of publishing all upper-air data.

Sources of Data

1. Daily Aerological Record of the Meteorological Office, London.
2. Daily Series, Synoptic Weather Maps, Part II, US Department of Commerce, Washington, D.C.
3. Daily Upper Air Bulletin, US Fleet Weather Central, Washington, D.C.
4. Täglicher Wetterbericht des Deutschen Wetterdienstes, Bad Kissingen.

REFERENCES

Bannon, J. K., 1954: Note on the structure of the high altitude strongwind belt in the Middle East in winter. *Quart. J. Roy. Met. Soc.* **80**, 344, pp. 218—221.

Bolin, B., 1950: On the influence of the earth's orography on the general character of the westerlies. *Tellus* **2**, 3, pp. 184—195.

Cressman, George P., 1950: Variations in the structure of the upper westerlies. *J. Meteor.* **7**, pp. 39—47.

Hess, Seymour L., 1948: Some new mean meridional cross sections through the atmosphere. *J. Meteor.* **5**, pp. 293—300.

James, R. W., 1951: A February cross-section along the Greenwich meridian. *Meteor. Mag.*, London, 954, **80**, pp. 341.

Mintz, Y., and Dean, G., 1951: Investigation of the general circulation of the atmosphere. Part II, Report 7. Univ. of California, Los Angeles.

Mohri, K., 1953: On the fields of wind and temperature over Japan and adjacent waters during winter o 1950—1951. *Tellus* **5**, 3, pp. 340—358.

Namias, J., and Clapp, P. F., 1951: Observational studies of general circulation patterns. *Compendium of Meteorology*, pp. 551—567.

Palmén, E., 1934: Über die Temperaturverteilung in der Stratosphäre und ihren Einfluss auf die Dynamik des Wetters. *Meteor. Z.*, **51**, pp. 17—23.

Rex, Daniel F., 1950: Blocking action in the middle troposphere and its effect upon regional climate. Part II. *Tellus* **2**, 4, pp. 275—301.

Riehl, Herbert, 1954: *Tropical Meteorology*. McGraw-Hill, New York.

Vuorela, L. A., 1948: Contribution to the aerology of the tropical Atlantic. *J. Meteor.* **5**, pp. 115—117.

Yeh, T.-C., 1950: The circulation of the high troposphere over China in the winter of 1945—1946. *Tellus* **2**, 3, p. 173—183.

On the Chemical Climate and its Variation with the Atmospheric Circulation Pattern

By

C.-G. ROSSBY, University of Stockholm and the Woods Hole Oceanographic Institution, and
H. EGNÉR, Royal Agricultural College, Uppsala.*

(Manuscript received January 15, 1955)

A study of the chemical composition of monthly precipitation samples from a number of stations in Sweden brings out the existence of significant variations in the relative proportions of chlorides and sodium depending upon the general character of the prevailing circulation. In maritime westerly flows the weight ratio of chloride to sodium decreases eastward to values which lie far below the value characteristic of sea water. In precipitation falling from arctic or polar continental air masses the chloride component is almost completely absent. The highest amounts of chlorides relative to sodium are observed in precipitation from warm and most air masses reaching Sweden from the south or southeast.

Significant variations are observed also in the yearly means of the sodium and sulphur deposits depending upon the dominance of maritime or continental air currents.

A sketch map of the "average" chloride concentration in European precipitation prepared from miscellaneous older data lends support to the results of the analysis of the monthly data. The separation of the chemical components indicated by this study would seem to be of considerable geochemical interest since it would permit different maritime salts to be deposited in widely separated parts of the continents. It is not inconceivable that a detailed study of the association between different circulation patterns and the corresponding chemical climates might permit us to draw inferences concerning variations in the atmospheric circulation during post-glacial times, e.g. from the data on the chemical stratification in raised bogs now becoming available through the investigations by *Mattson* and *Koutler-Andersson*.

It is the specific purpose of this paper to show that the weight ratio of chloride to sodium ions determined through the analysis of monthly accumulated precipitation samples varies in a systematic manner from month to month, and that the value of this ratio closely reflects the trajectory and life history of the principal air mass from which or through which the month's precipitation is falling. The range over which this ratio varies extends from 0.0 to about 3.5, the average value for sea water being about 1.8 and for pure sodium chloride 1.5. The fluctuations here established clearly indicate that Cl^- and Na^+ are not inseparable but rather that they, through various processes about which our knowledge is as yet mainly conjectural, may be detached from each other in leaving the surface of the sea or shortly thereafter. Nuclei containing sodium chloride and then certain other sodium salts appear to be precipitated at the most rapid rate while other chlorides seem to be capable of travelling long distances from their oceanic source regions without precipitation.

The more general purpose of our paper, however, is to call attention to the wealth of fascinating problems created by the availability, since November 1, 1954, of synoptic data (monthly means) on the chemical composition of precipitation and of the surface air at a large number of stations in Denmark, Finland, Norway and Sweden. The data for November and December 1954 are published as an appendix to this issue of Tellus. It is our intention to publish the data for subsequent months in quarterly installments.

The data available for the present study consist of the following:

a) The amounts of the principal chemical constituents contained in monthly precipitation

* The authors are indebted to the Knut and Alice Wallenberg Foundation for a generous grant which has enabled them to initiate the organization of an international network of chemical sampling points, and to the U. S. Office of Naval Research and the U. S. Weather Bureau which through their unfailing support have made it possible for the Institute of Meteorology to take up chemical meteorology as a major field of research. Contribution 770 from the Woods Hole Oceanographic Inst.

samples collected at a network of about 28 Swedish stations, operating from October 1, 1948, until September 30, 1950. A map showing the names and locations of these stations is given in a recent paper in Tellus (EMANUELSSON, ERIKSSON and EGNÉR 1954). Unfortunately, data on sulphur and magnesium are not available for this period.

b) Data on the amounts of various chemical components, including sulphur and magnesium, contained in monthly precipitation samples collected at seven stations in southern Sweden during the period July 1, 1951, until June 30, 1953. These stations were Alnarp, Bräkne-Hoby, Plönninge (Halmstad), Flahult, Bjärka-Säby, Skara (Lanna) and Ultuna. At these stations also air samples were collected monthly by means of continuously operating low capacity pumps (about 30 m^3 per month) and analyzed for the mean monthly concentration of the same elements (exclusive of NO_3—N) as those determined in the precipitation samples.

c) Data from an expanded net of Swedish stations at which both air and precipitation samples have been collected monthly since July 1, 1953. On November 1, 1954, this network was enlarged to include additional stations in Sweden, Norway, Finland and Denmark. The locations of these stations are given in the station map accompanying the chemical data appendix included in the present issue of Tellus. Since July 1, 1953, the hydrogen ion concentration (pH) and the electric conductivity of the monthly precipitation samples have been determined at all stations.

The chemical analyses were carried out at the Royal Agricultural College of Uppsala (Ultuna) under the supervision of one of us (Egnér). The meteorological analyses and the interpretation of the results were carried out at the University of Stockholm. The chemical technique now in use is described in the chemical data appendix.

In the report by Emanuelsson, Eriksson and Egnér, it was shown that during the two-year period October 1, 1948, to September 30, 1950, the weight ratio of the total amount of chloride to the total amount of sodium deposited per hectar and year decreased from a value of about 1.2 on the Skagerrack coast to about 0.6 in the vicinity of Stockholm—Ultuna. The separation of these two constituents was first brought out by CAUER (1949) in his studies of atmospheric condensation nuclei. Cauer states that the sodium and chloride ions added to the atmosphere from the sea through spray or bubble formation go through entirely different cycles. He believes that the chloride ions contained in the sea spray droplets, under the influence of the small amounts of ozone present in the lowest strata of the atmosphere (0—100 μg per m^3), may be oxidized to free chlorine gas which again, in the presence of sunlight and water vapour, is changed into hydrochloric acid. The escaping chlorine would be replaced in the initial droplets by bi-carbonate ions. Since the hydrochloric acid may be transported by the winds as a gas or may form droplets too small to serve as effective condensation nuclei, it is clear that the Cauer process would explain why rains falling within relatively short distances of the coast are characterized by relatively low and landward decreasing values of the weight ratio of chloride to sodium ions. Cauer's hypothesis is now being studied experimentally and theoretically by our collaborators and these studies will be reported in due time. It appears, however, that the effect observed by Cauer and verified by our precipitation data, might be accounted for by one or several alternative processes. KÖHLER and BÅTH (1952) have carried out an interesting experimental study of the separation of sea salt components, by blowing an air stream over a sample of sea water. Droplets of different size were analyzed separately for the main components. However, the maximum separation indicated by these experiments is insufficient to account for the observed geographical variation in Cl/Na.

The Cauer hypothesis, or any equivalent mechanism, carries with it one extremely important implication, apparently overlooked by Cauer but of considerable geochemical interest:

Presumably the proportions of chloride and sodium ions which leave the sea surface are the same as in the water but apparently some sodium is precipitated in combination with other negative ions than chloride in the coastal areas of western Europe. There must then develop an excess of chloride ions or of molecular chlorine in the free atmosphere. Unless one assumes that molecular chlorine is taken up directly by the ground or that

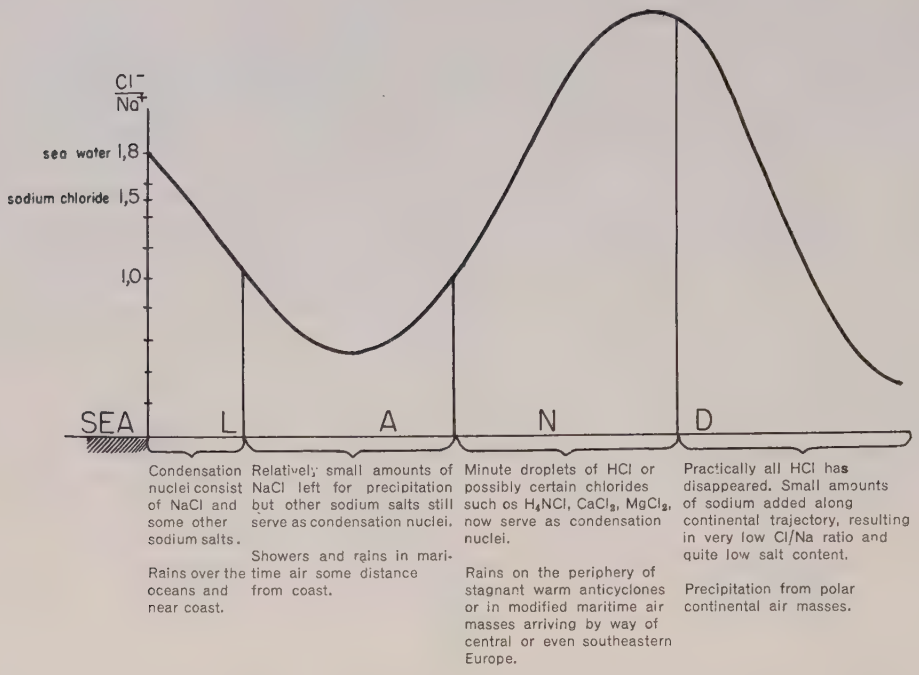

Cl⁻/Na⁺
sea water 1,8
sodium chloride 1,5
1,0
SEA
L
A
N
D
Condensation nuclei consist of NaCl and some other sodium salts.
Rains over the oceans and near coast.
Relatively small amounts of NaCl left for precipitation but other sodium salts still serve as condensation nuclei.
Showers and rains in maritime air some distance from coast.
Minute droplets of HCl or possibly certain chlorides such os H_4NCl, $CaCl_2$, $MgCl_2$, now serve as condensation nuclei.
Rains on the periphery of stagnant warm anticyclones or in modified maritime air masses arriving by way of central or even southeastern Europe.
Practically all HCl has disappeared. Small amounts of sodium added along continental trajectory, resulting in very low Cl/Na ratio and quite low salt content.
Precipitation from polar continental air masses.

Fig. 1. Schematic representation of the variations in the weight ratio of chloride to sodium in precipitation from an air mass moving in over the continent.

chlorates or hypochlorites not now identified in our precipitation samples maintain the chlorine mass balance in the atmosphere, one is forced to conclude that there must exist continental regions far removed from the oceanic source regions, in which the weight ratio of chloride to sodium in precipitation must be very much in excess of the value characteristic of sea water or of coastal precipitation. If the chlorides (exclusive of sodium chloride) only slowly come into play as condensation nuclei it would in fact be possible for rains in such interior regions to reach a higher absolute chloride concentration than rains falling closer to the coast. It was this paradox that provided the starting point for the present study.

We have attempted to sketch the variation of the chloride to sodium ratio in precipitation as a function of distance from the oceanic source region in Fig. 1. Leaving out the many important nuclei not containing either sodium or chlorides one would have to assume that the precipitation falling from air masses near the coast contain a mixture of sodium chloride and some other sodium salt. As the air travels further away from the coast most of the sodium chloride would be used up and ultimately the rain drops would contain mainly sodium salts other than sodium chloride. Finally, when also these sodium salts have been consumed, the minute hydrochloric acid droplets or corresponding chlorides would come into play as condensation nuclei. When this happens the chloride to sodium ratio would, of course, rise far in excess of the 1.8 value characteristic of sea water. If the air continues along a continental trajectory without access to chloride sources but is able to pick up small amounts of sodium, it is probable that the chloride to sodium ratio of the precipitation would again decrease.

Figures 2, 3, 4 and 5 represent an attempt to verify the corollary formulated above through a study of the geographic distribution of the chloride to sodium ratio during four selected months of dissimilar circulation patterns. Each figure contains the mean sea level pressure pattern over northern and central Europe for the month in question. The monthly precipitation values in mm at the chemical stations have been entered on each map. In each figure the

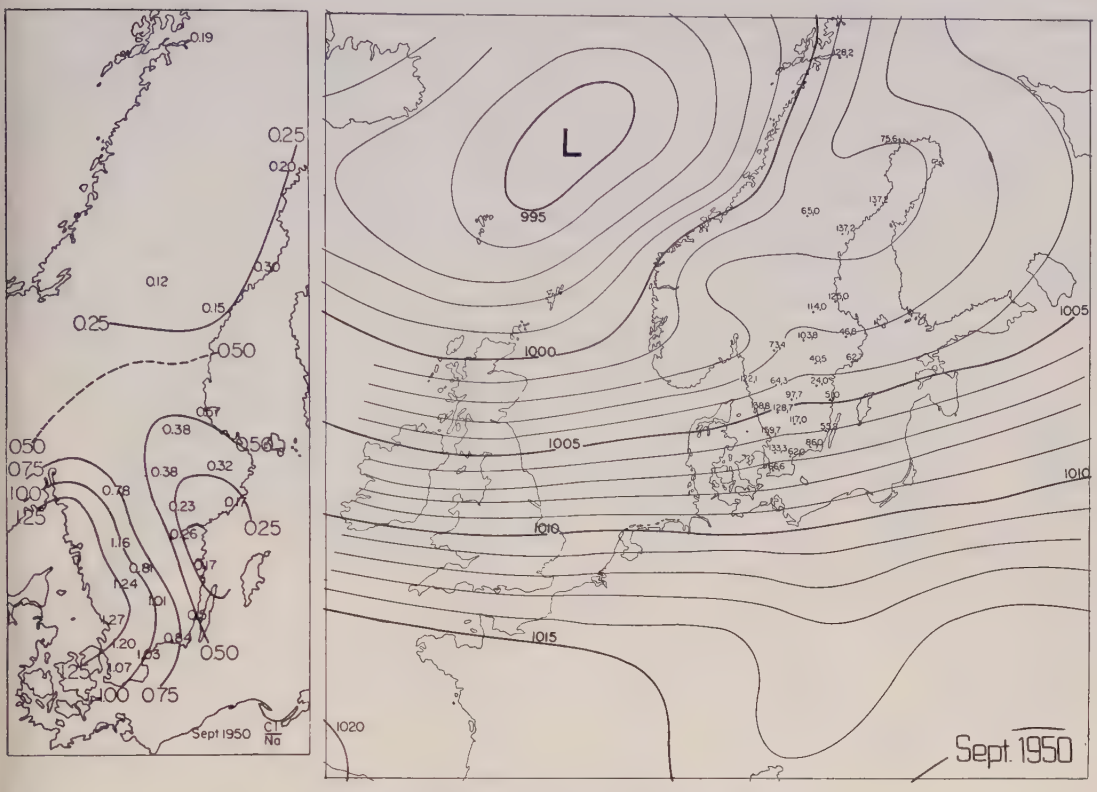


Fig. 2. Mean circulation pattern at sea level over Northwestern Europe during September 1950. The precipitation in mm at the chemical sampling points is entered on the pressure chart. The inset map gives the geographic distribution over Sweden of the weight ratio of chloride to sodium in the accumulated precipitation samples.

inset map gives the geographic distribution over Sweden of the weight ratio of chloride to sodium, expressed through isolines. It should not be necessary to stress that a monthly mean pressure pattern must be used with a great deal of caution in drawing conclusions concerning the paths of the rain-bearing winds; it has been suggested that mean charts should be constructed only for those days on which precipitation fell. However, the number of days with precipitation at any one station in Sweden is quite large and when twenty or thirty stations are considered such a map would become practically identical with the monthly mean chart. Figure 2 shows that in September 1950 straight westerly winds dominated the flow pattern in southern Sweden up to about the latitude of Stockholm. During this month the chloride to sodium ratio fell from about 1.3 on the Kattegatt coast to a minimum of 0.17, less than one tenth of the sea water value, on the Baltic coast. The precipitation decreased markedly from west to east.

December 1949 (fig. 3) was less uniform in its circulation characteristics but by and large it was dominated by strong west winds over southern Sweden. Also in this case the precipitation values decreased from Kattegatt to the Baltic coast and the chloride to sodium ratio decreased eastward, though less rapidly than in September 1950. In both cases this ratio reached very low values in northern Sweden, probably because the air reached that part of the country after a fairly long over-land journey and after considerable lifting over the mountains in Norway. Most of the sodium chloride must have been precipitated by the time the rain-bearing winds reached northern Sweden, other sodium salts than NaCl having taken over the role as condensation nuclei.

May 1950 (fig. 4) represents a month with a rather unusual circulation pattern, northerly

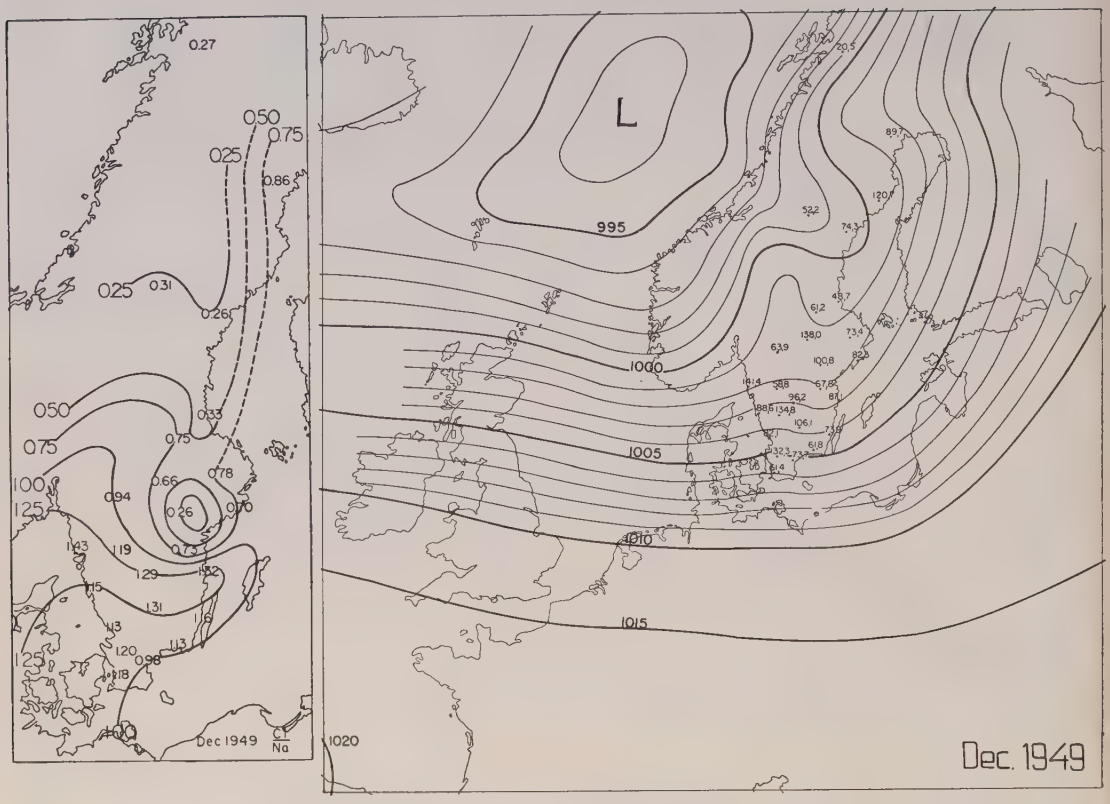


Fig. 3. Circulation pattern and weight ratio of chloride to sodium during December 1949. See legend to Fig. 2 for details.

or even northeasterly winds over the entire country. The precipitation amounts were low but sufficient for analysis. Amazingly low values of the chloride to sodium ratio are observed, less than 0.3 at all stations but one. In view of the hypothesis proposed by Cauer it should be pointed out that northerly winds during the spring presumably are characterized by relatively high ozone concentrations in the lower air strata.

November 1949 (fig. 5) was characterized by persistent advection of warm and moist air from the Mediterranean and Balkan region, air masses of Atlantic origin having made a long sweep eastward over southern Europe. The rainfall in southeastern Sweden during this month was profuse, up to two or three times the normal value. In these rains the chloride to sodium ratio reached values of up to 3.5 along the southern and southeastern coast of Sweden. The absolute amounts of chloride deposited in those districts were relatively large, in excess of 1 kg per hectar (10^4 m^2) at several stations and reaching 2.06 kg per hectar at Tovehult.

The four cases presented above strongly support the proposition that air masses leaving the sea first lose sodium chloride through precipitation, then certain other sodium salts, but that the remaining chlorides, which are not associated with sodium may stay with the moist air masses until long after the major portion of the sodium has been precipitated.

The extremely low chloride ratios observed in May 1950 remain to be explained. It is conceivable that the slight total precipitation observed during that month fell from air masses with a very long continental history, air masses which had moved around the blocked low pressure system located in the White Sea region. In the course of this long continental history, the air would ultimately

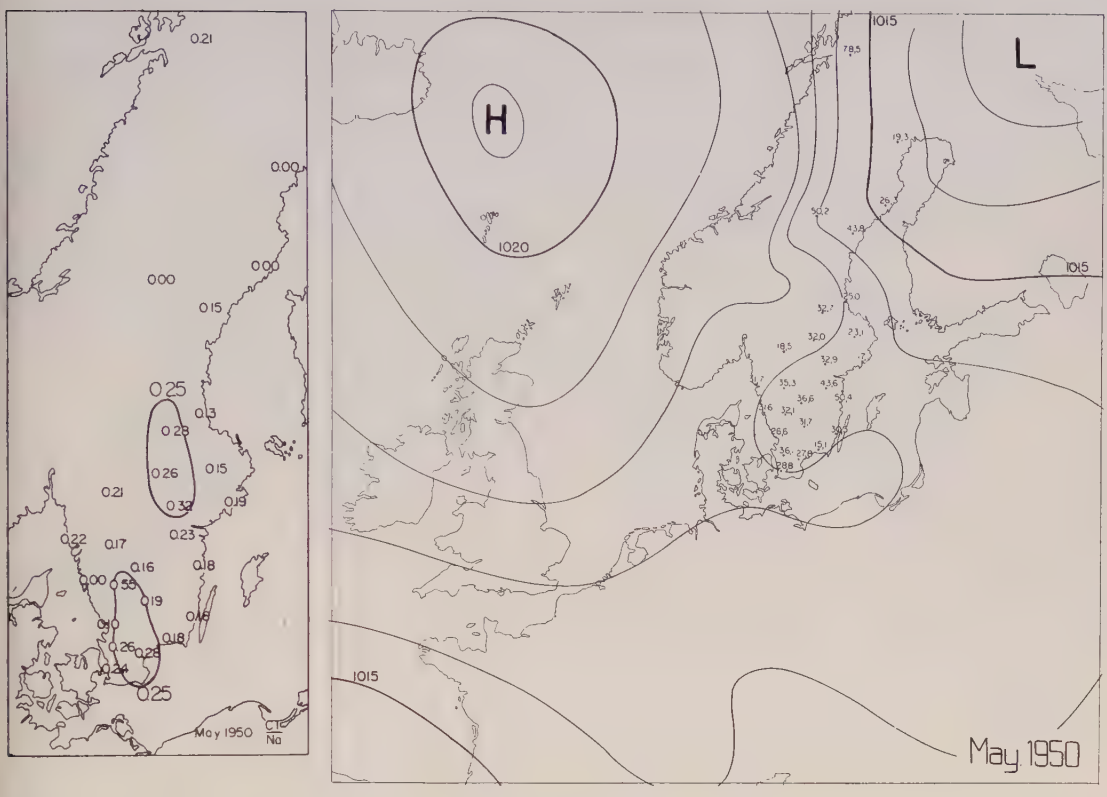


Fig. 4. Circulation pattern and weight ratio of chloride to sodium during May 1950. See legend to Fig. 2 for details.

lose the hydrochloric acid particles or chlorides through precipitation. Since there are no continental sources of chlorides, with the exception of industrial pollution and salt deserts, sources which hardly would have to be considered in these high latitudes, while there always would exist some opportunity for the air to pick up small amounts of sodium together with other typical continental cations, it is conceivable that the low chloride ratio in the precipitation of the northeasterly winds might be explained as associated with a second minimum in the chloride ratio curve, a minimum reached only after most of the chlorides have been precipitated without replacement. This second minimum would evidently be associated with very low salt concentration in the precipitation. To study this problem further we have looked into the precipitation chloride ratio for the month of April 1954. During this month, northerly to northeasterly winds prevailed over northern and central Sweden, northwesterly winds on the west coast of Sweden and in the southernmost parts of the country (fig. 6). It is probable that this last current to a large extent was fed by air from Central Europe brought towards southern Scandinavia in a broad sweep out of a stagnant anticyclone over the British Isles. The inset map shows the distribution of chlorides in mg per liter. On the same chart the lower number at each station gives the chloride to sodium ratio. The contrast between the low chloride concentration in the precipitation from the continental northeasterly current and the extremely high concentration in the precipitation over southern Sweden is striking. The high chloride ratio values in southern Sweden suggest that much of the pure sodium chloride particles already have been removed from the air and that other chlorides to a large extent have taken their place.

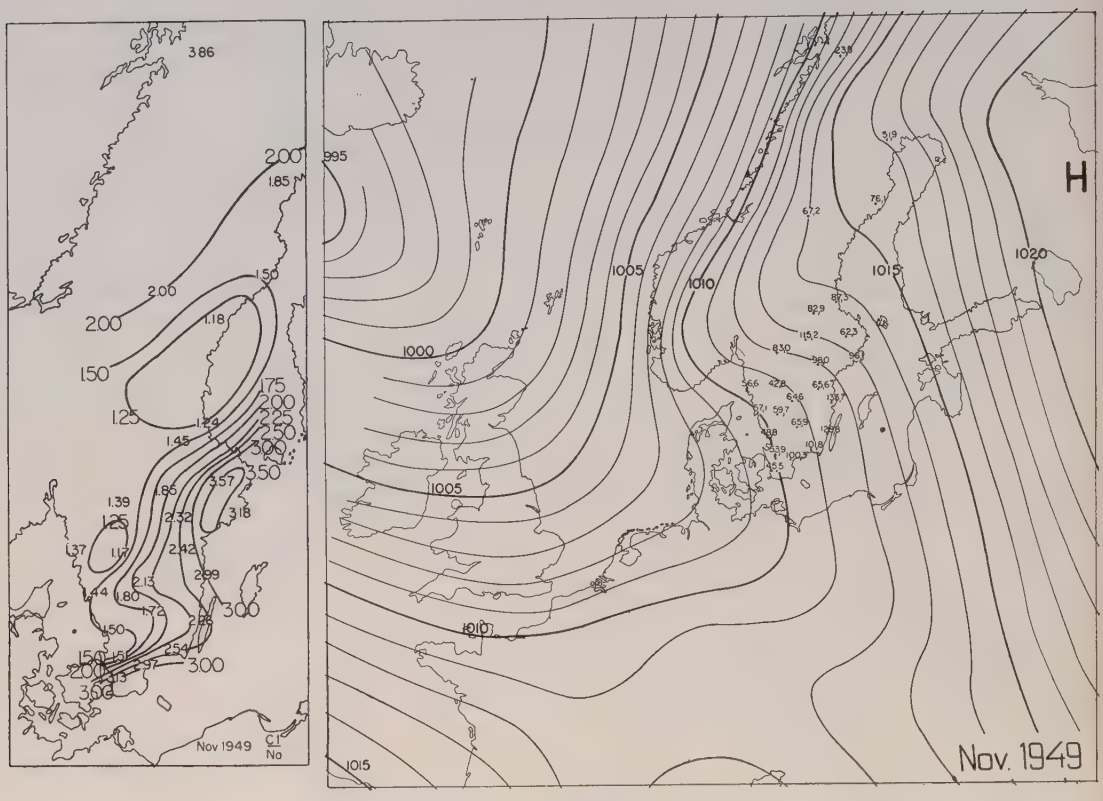


Fig. 5. Circulation pattern and weight ratio of chloride to sodium during November 1949. See legend to Fig. 2 for details.

We have attempted to make use of the combined precipitation and air analyses conducted since July 1, 1953, to study the chloride content in the surface air of currents reaching Sweden from the south. As a first step, we selected eleven stations (fig. 7) in the eastern half of Sweden and plotted for each one of these stations the monthly mean chloride content expressed in micrograms per m^3 of air as a function of time (fig. 8). It is striking that all these curves are characterized by a well-defined maximum in December 1953. We then constructed (fig. 9) curves representing the mean concentration at these eleven stations of chloride, sulphur, sodium and magnesium as functions of time. The sodium and magnesium values for December 1953 exhibit no peculiarities, but the sulphur shows a minimum. (March 1954, on the other hand, is characterized by an extremely high sulphur value.) In December 1953 the ratio of the average chloride concentration to the corresponding sodium concentration as given by the curves in figure 9 reached a value of 4.0. The corresponding ratio computed from the total precipitation deposits at the same eleven stations was 1.74. This suggests the presence in the lowest air layers of HCl in gaseous form or in the form of minute droplets too small to be collected effectively by the falling rain.

The dominant circulation pattern of December 1953 is illustrated by figure 10. It represents the mean sea level pressure pattern for the two-week period December 9—December 23, inclusive, during which southern and central Sweden were dominated by southerly or even southeasterly winds bringing in warm air from southern and central Europe. In the first and the last week of the month moist air was brought into southern Sweden from the southwest and west. The northernmost part of the country was under the intermittent influence of maritime air masses throughout the

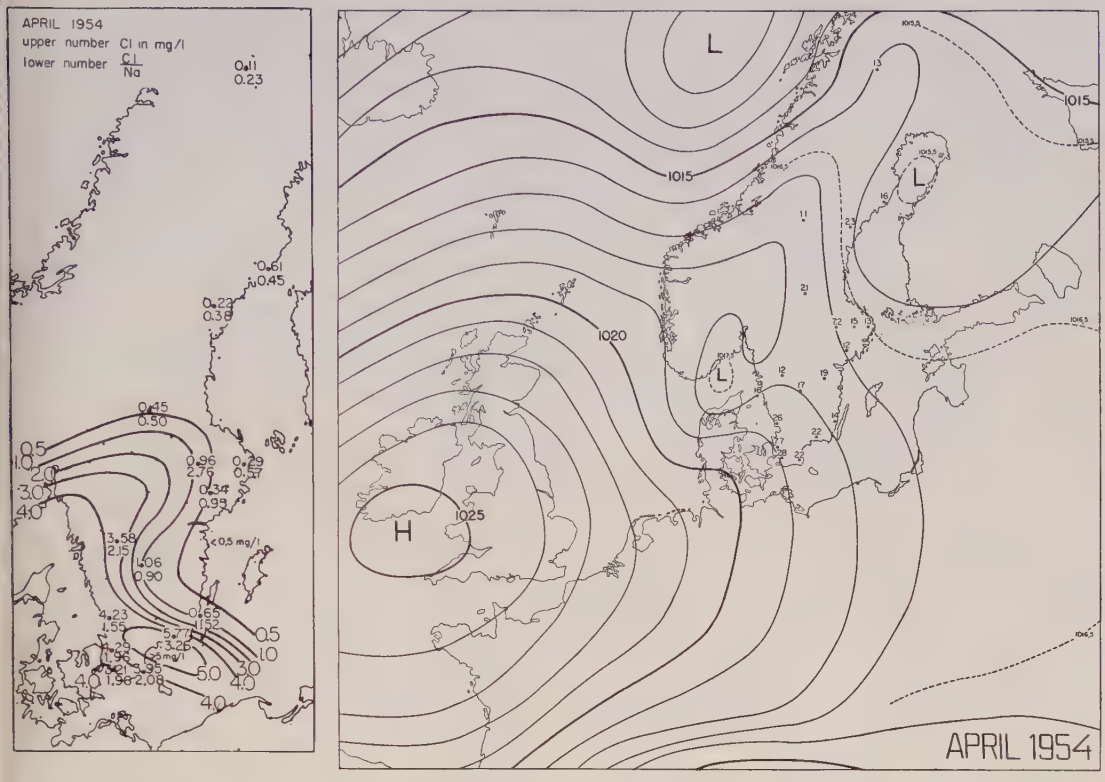


Fig. 6. Mean circulation pattern at sea level during April 1954, with precipitation amounts in mm at the chemical sampling points. The inset map gives the distribution of chlorides in the precipitation (mg pr liter) expressed through "isochlors". The lower figure at each sampling point gives the weight ratio of chloride to sodium.

major part of the month but received very small amounts of precipitation. The mean sea level pressure distribution for the entire month is not reproduced but it shows mean monthly pressures well above normal (10—11 mb in southern Sweden, 1—2 mb in northernmost Lappland). The flow was in the main anticyclonic and hence the precipitation was below normal practically everywhere except in the mountainous western part of northern Sweden.

The dominant southerly to southeasterly flow in southern Sweden during December 1953 was associated with a tongue of quite low pH-values in the precipitation samples (fig. 11). This observation supports the assumption that the prevailing air current contained droplets of hydrochloric acid. It is of some interest to compare this month with March 1954 (fig. 12), the month during which the concentration of sulphur (sulphate and sulphite ions) in the air samples from the eleven eastern stations reached a pronounced maximum. Also during this month southerly winds prevailed in southern and central Sweden, southerly to southwesterly winds in northern Sweden. In December 1953, temperatures were well above normal in southern Sweden, while March 1954 was characterized by about normal temperatures except in the north where they were in excess of normal. The March precipitation fell in the form of snow, except far south, the December precipitation mainly in the form of rain. One may perhaps infer that the December circulation pattern was dominated by air masses originally of maritime origin but having made a broad sweep over southern Europe. During March 1954 the country must at least at times have been under the influence of truly continental air masses from the interior of the Eurasian continent. The maximum of sulphur observed during this month may then have been the result, at least in part, of

Fig. 7. Location of eleven sampling stations utilized in the analysis of the month-to-month fluctuations in the chemical content of the air.

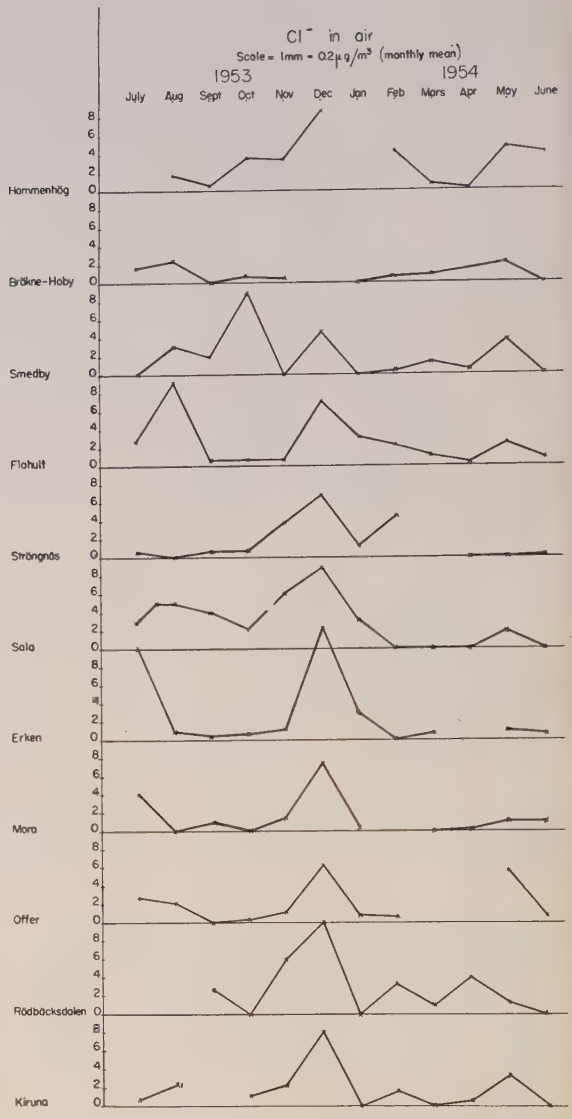


Fig. 8. Variation with time of the chloride content of the air (in μg per m^3) at eleven sampling points.

increased winter time combustion in central and eastern Europe.[1]

The study of these two months suggests that blocking of the "normal" west-wind circulation over northern Europe may be associated with marked changes in the chemical composition both of the precipitation and of the air. We have attempted to explore this possibility further by a study of changes in

[1] Mr. Claes Rooth of the Institute of Meteorology has looked into the air circulation over Europe in March 1954. His comments follow:

"The mean pressure chart for the month shows a well developed continental anticyclone together with a fairly deep cyclone over the north Atlantic. This indicates predominantly southerly flow over Scandinavia during the period. An inspection of the daily charts prepared by the Swedish weather service shows, as might be expected, a sequence of cyclonic and anticyclonic weather patterns over northern Europe. In central Europe, however, a weak anticyclonic circulation with wide areas of fog and low stratus dominated the month, preventing the vertical spreading of the pollution from the large industrial and urban areas. Extreme cases of such concentrations of highly polluted air under an inversion are known from Belgium, England and the United States where they have had even fatal consequences.

From the reservoir thus formed, contaminated air has apparently been brought up towards Scandinavia during the periods of SW—to SE flow. It has been stated by several workers that minute droplets of sulphuric acid are among the most persistent of the air pollutants that have been studied, which might explain partly why only the sulphur concentrations are abnormal in the air samples and partly why these high values are not reflected in the precipitation samples, which in fact show sulphur minima at several stations. In connection with the latter question it must be noted, however, that most of the precipitation fell during the first week of the month before the characteristic pattern described above had developed."

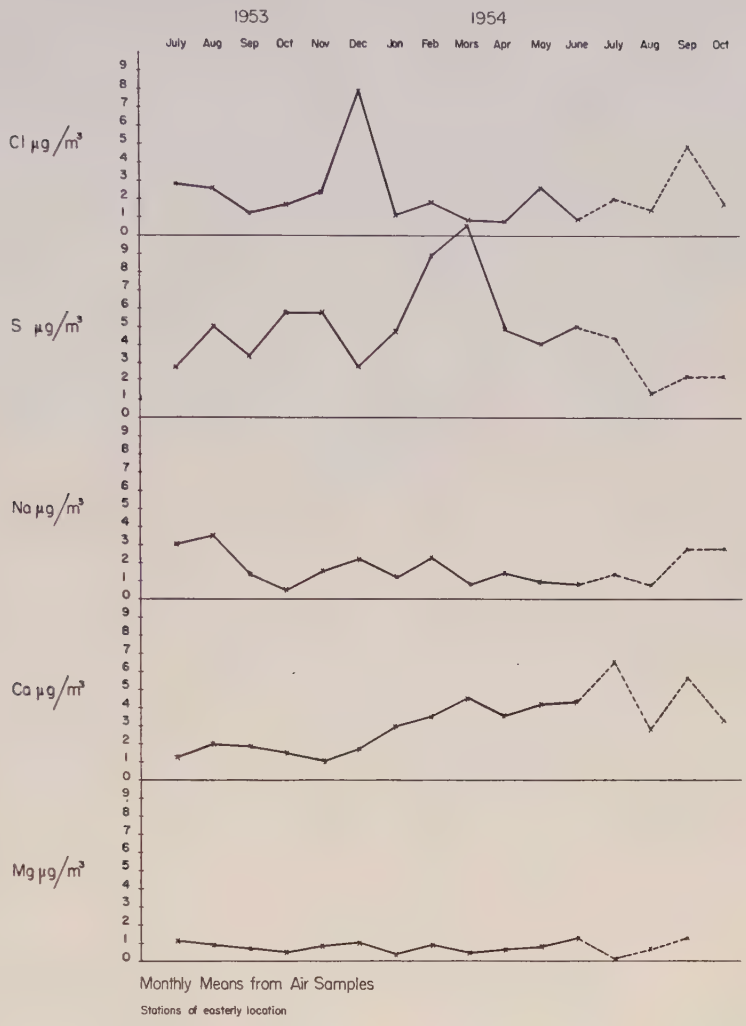


Fig. 9. Variation with time of the average concentration of chlorides, sulphur, sodium, calcium and magnesium (in μg per m^3 of air). Curves represented in this figure obtained by averaging the concentrations measured in he monthly samples from the eleven stations in Fig. 7.

the annual mean values of the various chemical components brought to the ground by precipitation and by a comparison of these changes with the changes in the corresponding mean annual circulation patterns. The data at our disposal are inadequate but they suggest nevertheless that the "ventilation" of southern Sweden by fresh maritime air masses decreased successively from 1949 until 1953 and that southern and southeastern Sweden during these years gradually became more exposed to air masses which had spent some time over central Europe and thus lost much of their original sodium content. To demonstrate this, it is perhaps sufficient to point out the trends of the annual deposits of various chemical components at a few representative stations situated away from the coast where data are available for several years and where local sources of pollution appear relatively unimportant:

	Sodium deposit in kg per har and year			
	1949	1952	1953	July 1, 1953—June 30, 1954
Skara	9.46	4.46	4.48	3.92
Flahult	9.20	3.74	3.65	4.32
Bräkne-Hoby	5.80	5.23	3.90	4.88

	Chloride deposit in kg per har and year			
	1949	1952	1953	July 1, 1953—June 30, 1954
Skara	9.37	7.28	7.00	6.61
Flahult	10.34	5.72	5.14	5.71
Bräkne-Hoby	8.03	7.59	6.30	8.54

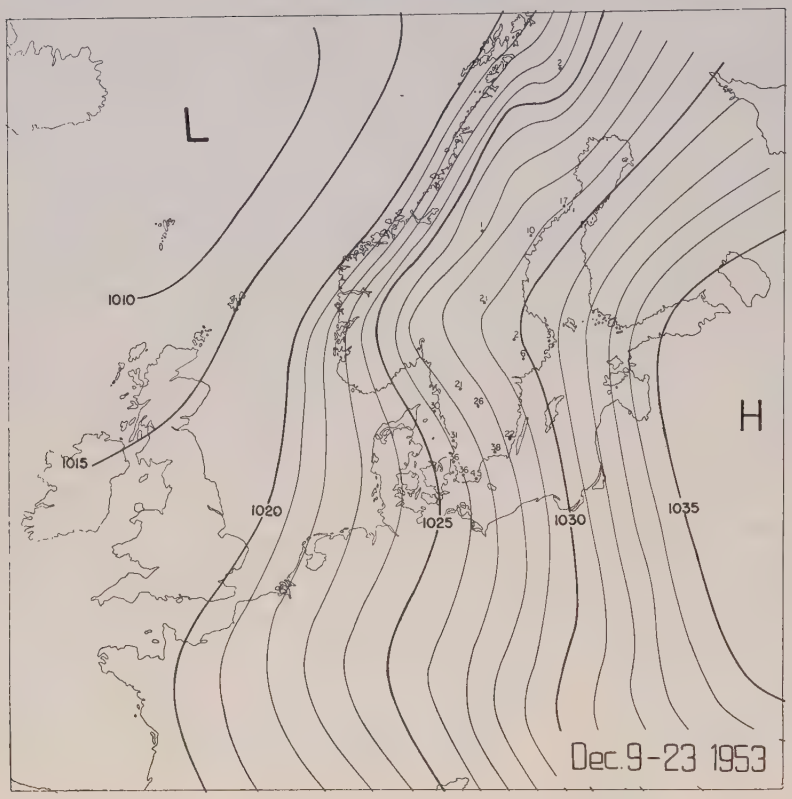


Fig. 10. Mean circulation pattern at sea level during the period December 9—23, inclusive, 1953.

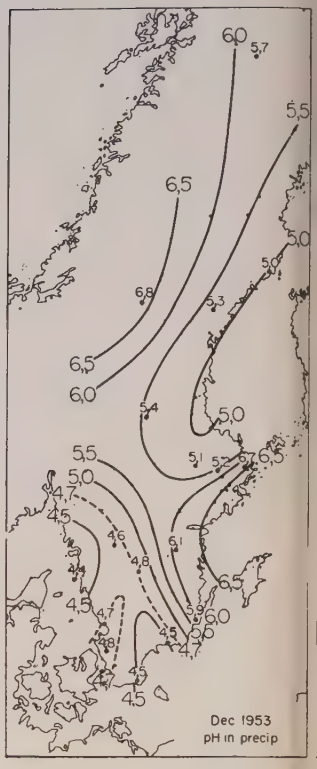


Fig. 11. Geographic distribution of the hydrogen-ion concentration in the accumulated precipitation for December 1953.

Ratio of chloride to sodium

	1949	1952	1953	1953—1954
Skara	0.99	1.63	1.56	1.69
Flahult	1.12	1.53	1.41	1.32
Bräkne-Hoby	1.05	1.45	1.62	1.75

Annual precipitation in mm

	1949	1952	1953	1953—1954
Skara.....	547	398	624	516
Flahult.....	773	611	705	739
Bräkne-Hoby	515	645	346	470

Data on the sulphur content are not available for the entire period and thus we have been forced to group the observations differently to bring out such trends as may exist. Hence, one finds:

12 month period	Annual deposit of sulphur in kg per hectar and year: July 1951—June 1952	July 1952—June 1953	July 1953—June 1954
Skara.....	6.10	6.12	5.94
Flahult	5.24	5.28	7.05
Bräkne-Hoby	6.37	7.47	8.81

Between 1949 and 1953 the sodium deposits fell off to one half or one third of their original value. The chloride deposits decreased too, but much less markedly. On the other hand, the two southeastern and eastern stations (Bräkne-Hoby and Flahult) showed rising deposits of sulphur. Taken together, the data point to an increasing frequency of air mass invasions from central Europe. To test this conclusion we have computed sea level mean pressure charts for the two years 1949 and 1952. (Figures 13 and 14.) The chart for 1949 shows a marked west wind circulation over the entire Scandinavian peninsula. In 1952, on the other hand, the most striking feature of the mean chart is a narrow high-pressure ridge which extends northward through the major part of Scandinavia. It follows that in 1952 the geostrophic winds across central Sweden must have blown almost as often from the

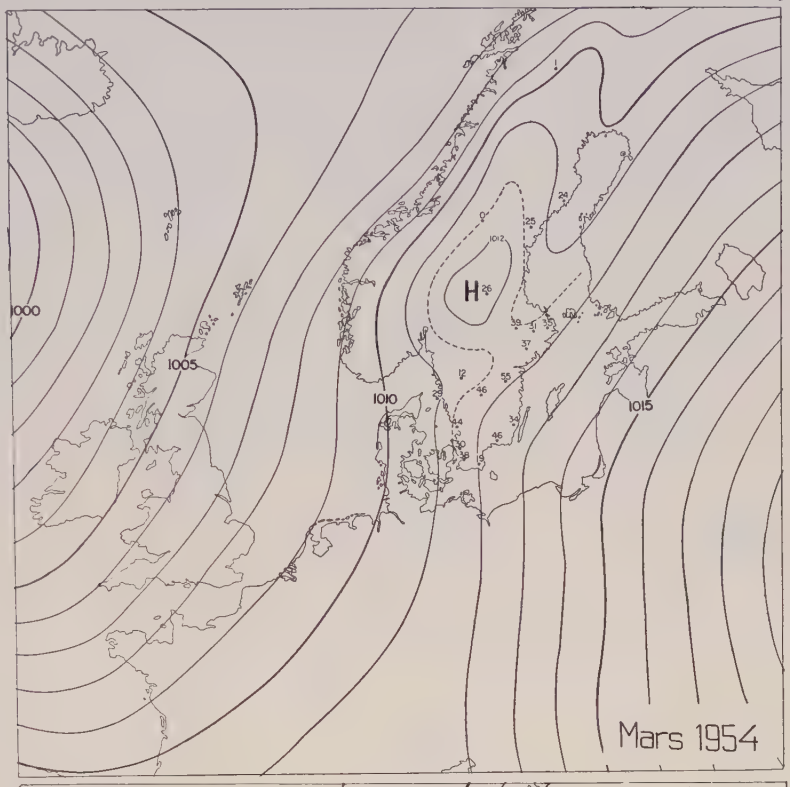


Fig. 12. Mean circulation pattern at sea level during March 1954.

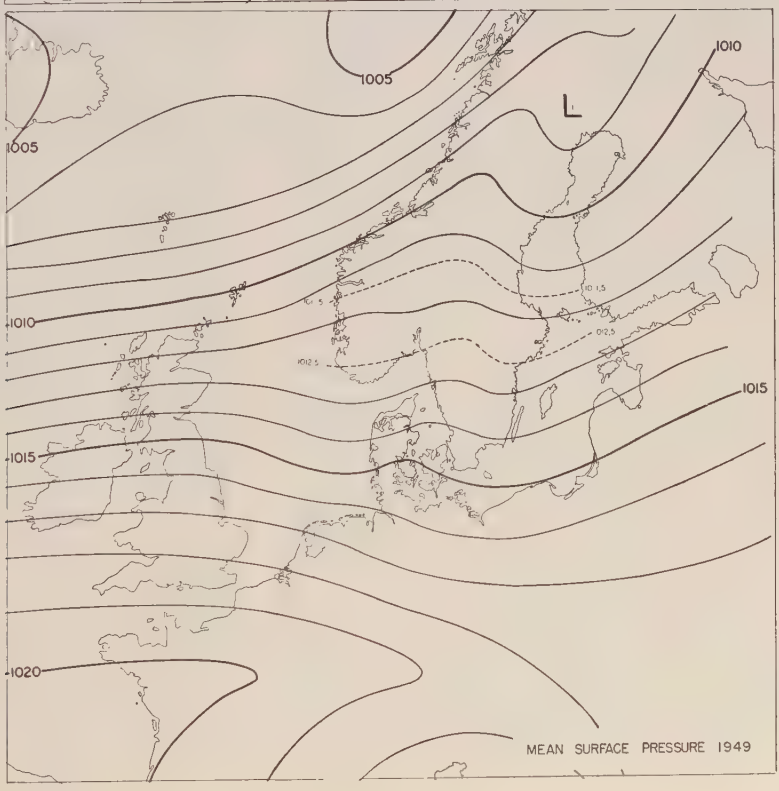


Fig. 13. Mean circulation pattern at sea level during the calendar year 1949.

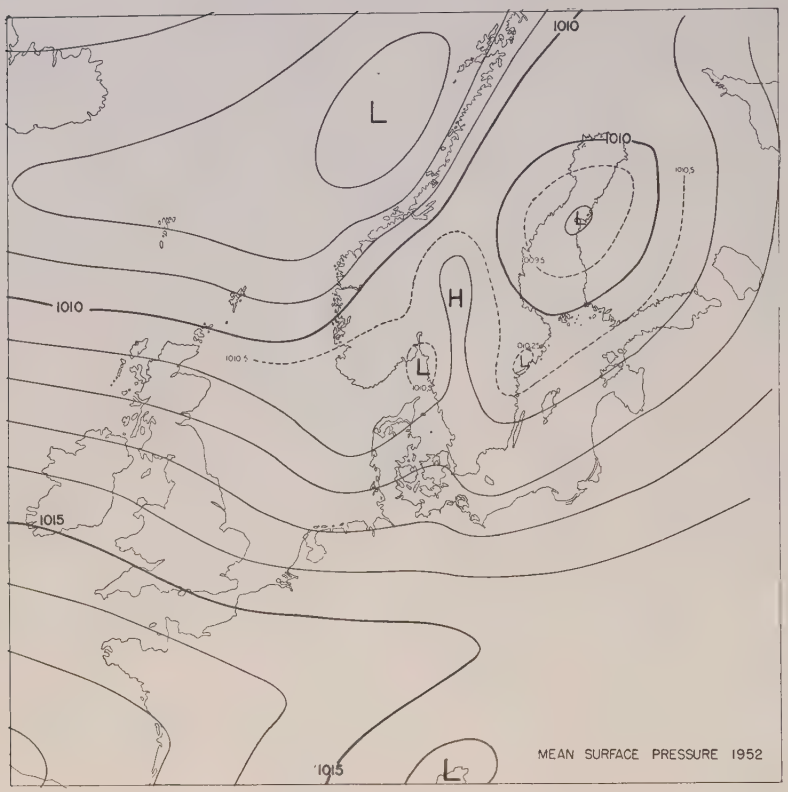


Fig. 14. Mean circulation pattern at sea level during the calendar year 1952.

east as from the west, *i.e.* the normal westwind "ventilation" of the peninsula had to a large extent been replaced by air currents from the interior of the continent. This point is brought out quite clearly by the graphical representation in fig. 15 of the frequency distribution of easterly and westerly geostrophic sea level winds across a north-south section in south-central Sweden.

The great variations from year to year demonstrated above show that the mean charts presented by EMANUELSSON, ERIKSSON and EGNÉR (loc. cit.) in no sense may be regarded as "normals" but merely as cartographic representations of the chemical composition of the precipitation in Sweden during one particular two-year period.

We shall finally attempt to check the various indications obtained from the study of our own data through an inspection of measurements in other parts of Europe. For this purpose we have attempted to sketch, in Fig. 16, the probable "average" distribution of chlorides in precipitation over northern and central Europa generally. The data for this sketch map have been obtained from the rather complete references contained in the survey by E. ERIKSSON (1952). A similar very useful tabulation of chloride data has been made by DRISCHEL (1940).

All chloride values entered on the sketch map in Fig. 16 have been expressed in mg per liter and it has been necessary to piece together data from periods widely separated in time. Observation series extending through less than one full year have not been used since the variations in the chloride content from month to month are extremely large. Likewise we have omitted many observation series from sampling points near the exposed western coasts of Europe, since the exact values of the high concentrations in the immediate vicinity of a coast clearly depend upon accidental details of orography and exposure. Valuable general information concerning the strong coastal gradients of the chloride concentra-

tion may be obtained from the investigations by FARCY (1931) and by LEEFLANG (1938).

The data from the British Isles (for instance RUSSEL and RICHARDS 1919) show a well-defined minimum in the interior of England (Cirencester 3.2 mg per liter, 26 years, Rothamsted 2.4 mg per liter, 27 years). Apart from the excellent coastal gradient data collected by FARCY (l. c.) the observations from France are inadequate but there are old series from Paris (one year, 1851) which gives a concentration of 2.3 mg per liter, in good agreement with the data from Cirencester and Rothamsted. A similar series from Nantes (about 90 km from the coast) for 1863 gives 8.4 mg per liter (this value is placed in the wrong column in table 10, p. 286 of ERIKSSONS survey (l. c.)).

In Germany data are available from several large industrial cities including Berlin, Dortmund and Essen. All these places show much higher values than the interior of England and represent undoubtedly the effects of strong local industrial pollution. There remains a one-year series from an apparently undisturbed sampling point at Müncheberg, approximately 50 km E of Berlin, with a concentration of 3.6 mg per liter (one year 1933—1934, LIESEGANG, 1934).

We have entered on the map the danish data (a one-year series for 1925—1926) obtained by HANSEN (1931). Likewise we have included a selection of the Swedish sampling points analyzed by EMANUELSSON, EGNÉR and ERIKSSON (1954). These data pertain to the calendar year 1949.

The data from the European part of the USSR are based on measurements made in 1909—1910 by WITUYNJ (1911). A statement by G. A. MAKSIMOVITJ (1953) in a lecture before the USSR Academy of Sciences seems to imply that these values still are accepted by USSR scientists as representative. Wituynj's measurements show notably high concentrations of chlorides in the Black Sea region decreasing northwards and towards Asia.

The most striking features in our sketch map are the existence of a chloride minimum extending from the interior of France towards eastern Sweden and the existence of a marked second chloride maximum in the eastern and southeastern part of Europe. This second maximum fits in well with the high chloride concentrations observed in SE Sweden during

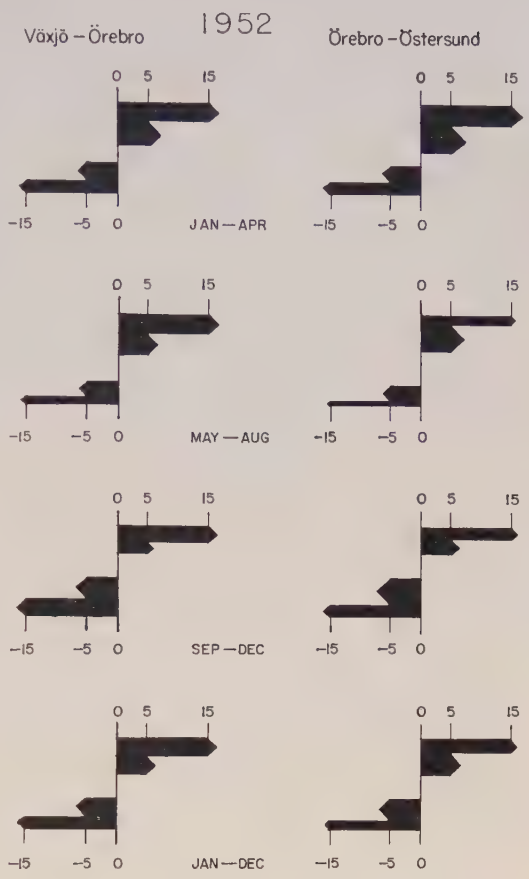


Fig. 15. Sea level frequency distributions for the year 1952 of geostrophic wind components from the west and from the east across two north—south sections in southern Sweden (Växjö—Örebro) and central Sweden (Örebro—Östersund). Thickness of arrows measures relative frequency of occurrence. Wind velocities expressed in mps.

November 1949 and in the air samples for December 1953. The explanation of this secondary maximum may lie in some process of ion separation operating on the sea salt nuclei but one can not a priori exclude as insignificant such other effects as volcanic and other continental sources of chlorides.

The separation of the individual chemical components of the nuclei brought out by some of the monthly patterns analyzed above would seem to fit in well with the results by JUNGE (1952, 1954) and WOODCOCK (1952) concerning the rôle of the giant nuclei in the precipitation mechanism. According to these authors the

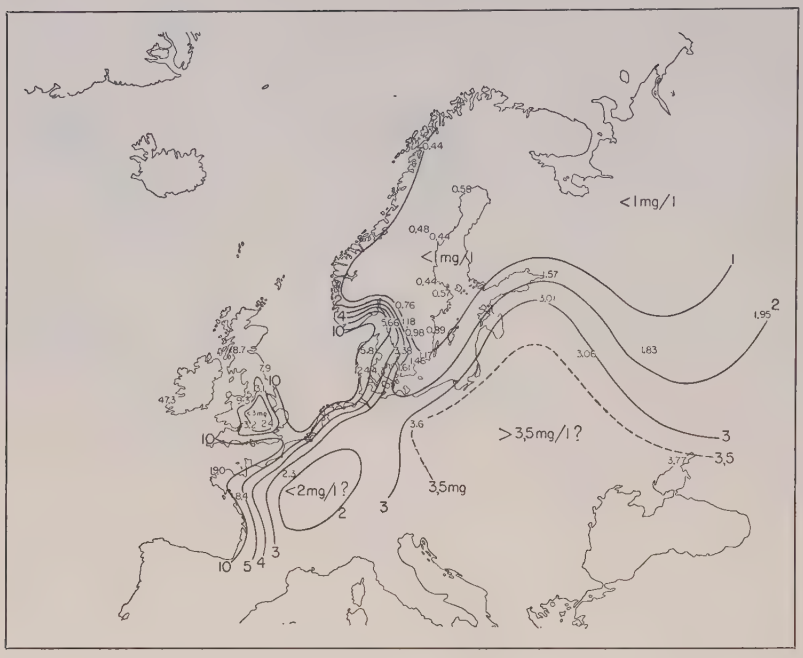


Fig. 16. A preliminary map of average concentration of chlorides (in mg per liter) in precipitation over Europe. Analysis based on records from different epochs and of very different length. Note proposed existence of minimum extending from central France towards eastern Sweden and of a secondary maximum over southern Russia.

largest nuclei consist mainly of sodium chloride and it is these nuclei which most readily lead to the formation of cloud droplets of precipitable size. One should therefore expect precipitation at stations not too far removed from the coast to possess a characteristic chloride over sodium ratio of about 1.5. After the precipitation of the sodium chloride particles there would remain in the air, or in the clouds, an excess of chlorides, perhaps in the form of H_4NCl or $MgCl_2$. The fact that the Cl/Na ratio even in straight westerly flows may drop to values considerably below 1.5 suggests however, that other sodium-containing nuclei become active before the chloride nuclei take over the precipitation-forming mechanim. With Dr JUNGE's permission we include a statement in a recent letter (February 20, 1955) to one of us (C.-G.R.):

"... Es waren mir zwar schon, auch von eigenen Untersuchungen, starke Abweichungen des Cl/Na-Wertes von dem des Seewassers bekannt, dass sie aber so stark sind und so augenscheinlich mit der Grosszirkulation verbunden sind, war doch eine Überraschung. Wenn ich Ihre Ergebnisse mit den meinen vergleiche, kommt mir der Verdacht, dass wir möglicherweise drei Komponenten unterscheiden müssen, die zum Gehalt des Regenwassers beitragen: Die Seesalzkomponente, die im wesentlichen 1.8 liefert, eine Gaskomponente, die eine Erhöhung dieses Wertes liefern würde und die ich jedenfalls in südlichen Breiten einwandfrei nachweisen konnte und drittens eine mineralische Staubkomponente, die zusätzliches Na liefert und das Verhältnis unter 1.8 sinken lässt.

"Das gefundene Verhältnis würde dann eine Mischung dieser drei Komponenten darstellen und natürlich von der Grosszirkulation abhängen. Nun ist die Lage Schwedens offenbar so, dass selbst bei überwiegenden Westwinden keineswegs mit einer reinen Seesalzkomponente zu rechnen ist. Den starken Anstieg der Werte bei südlicher Luftzufuhr würde ich als das Auftauchen der Gaskomponente deuten, möglicherweise unterstützt von industriellen Verunreinigungen. Das starke Absinken bei östlichen und nördlichen Winden wäre das Überwiegen der mineralischen Staubkomponent die ebenfalls noch bei westlichen Winden von Einfluss ist."

Regardless of the nature of the mechanism or mechanisms responsible for the separation of the individual chemical components observed in precipitation or in the air, it is quite evident from the data presented above that the chemical climate varies in a striking

manner with changes in the circulation patterns. In certain respects the association of a specific chemical climate with a specific circulation type may to some extent depend on man-made sources of pollution. By and large, however, one should expect the association to be stable and applicable also to past climates. The relative fluctuations in the temperature and precipitation climates with changes in circulation pattern are, at least in Northwestern Europe, small compared with the fluctuations in the chemical climate. There might therefore exist a possibility to trace post-glacial climatic fluctuations through their chemical consequences. In a recent major investigation MATTSON and COUTLER-ANDERSSON (1954) have made a detailed study of the chemical composition of different strata in a raised bog in SW Sweden. They find striking variations in the relative proportions of various elements and in the pH. It is tempting to try to associate these variations with variations in the circulation pattern, to check conclusions concerning post-glacial climatic fluctuations which have been drawn previously mainly from pollen-analytical studies. Such an attempt would, however, require a detailed investigation of the association between the circulation pattern and the deposition through precipitation of the bivalent ions, for instance Ca^{++} and Mg^{++}, which are most closely bound by the humic acids, and which therefore are capable of exhibiting a measurable stratification.

Acknowledgement

In the preparation of this paper numerous members of the Institute of Meteorology have been of great assistance. The various sea level mean maps have been plotted and analyzed by Mr P. Bergthorsson of the Icelandic Weather Service, and by mr E. Edström of the Institute staff. The frequency analysis of west and east wind components and certain other calculations referred to in the text was carried out by Mr Cl. Rooth of the Institute staff. The pH-chart for December 1953 was prepared by mr Earl Barrett, of the Woods Hole Oceanographic Institution. The authors have benefitted from many fruitful consultations with these persons, and with Mr. E. Eriksson of the Institute staff. We are particularly indebted to Mr F. Ludlam of Imperial College, London, who throughout the autumn of 1954 in a most invigorating manner participated in and contributed to the discussion of our problem and to Dr. Ch. Junge, with whom we have maintained a steady correspondence.

REFERENCES

CAUER, H., 1949: Ergebnisse chemisch-meteorologischer Forschung, Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B, Band I, pp. 221—256.

DRISCHEL, H., 1940: Chlorid-, Sulfat- und Nitratgehalt der atmosphärischen Niederschläge in Bad Reinerz und Oberschreiberhau im Vergleich zu bisher bekannten Werten anderer Orte, Der Balneologe, 7. Jahrgang, Heft 11, pp. 321—334.

ERIKSSON, E., 1952: Composition of Atmospheric Precipitation, II, Tellus, Vol. IV, No. 4, pp. 280—303.

EGNÉR, H., and ERIKSSON, E., 1955: Current Data on the Chemical Composition of Air and Precipitation, Tellus, Vol. VII, No. 1, pp. 134—139.

EMANUELSSON, A., ERIKSSON, E., and EGNÉR, H., 1954: Composition of Atmospheric Precipitation in Sweden. Tellus, Vol. VI, No. 3, pp. 261—267.

FARCY, L., 1931: La composition des eaux de pluie de la région de Brest et du Nord finistère. Ann. Hyg., **9**, p. 504.

HANSEN, F., 1931: Undersøgelser af regnvand. Tidskr. f. Planteavl., **37**, pp. 123—150.

JUNGE, CHR., 1953: Die Rolle der Aerosole und der gasförmigen Beimengungen der Luft um Spurenstoffhaushalt der Troposphäre. Tellus, Vol. V, No. 1, pp. 1—26.

— 1954: The chemical composition of atmospheric aerosols I: Measurements at Round Hill Field Station, June—July 1953. Journal of Meteorology, Vol. XI, No. 4, pp. 323—333.

KÖHLER, H., and BÅTH, M., 1952: Quantitative Chemical Analysis of Condensation Nuclei from Sea Water. Nova Acta Regiae Societatis Scientiarium Upsaliensis, Ser. IV, No. 7, pp. 1—24.

LEEFLANG, K. W. H., 1938: De chemische Samenstelling van den Neerslag in Nederland. Chem. Weekbl. **35**, pp. 658—664.

LIESEGANG, W., 1934: Untersuchungen über die Mengen der in Niederschlagswässern enthaltenen Verunreinigungen. Kleine Mitt. Mitgl. Ver. Wasser-, Boden- und Lufthygiene, **9**, pp. 350—355.

MAKSIMOVITJ, G. A., 1953: Atmospheric Precipitation and its rôle as carrier of dissolved substances (in Russian). Akademija Nauk SSSR, Doklady N S 92, p. 401.

MATTSON, S., KOUTLER-ANDERSSON, E., 1954: Geochemistry of a raised bog. Ann. Royal Agr. College, Sweden, 21, pp. 321—366.

RUSSEL, E. J., and RICHARDS, E. H., 1919: The amounts and composition of rain and snow falling at Rothamsted. J. Ag. Sci., **9**, pp. 309—337.

WITUYNJ, J., 1911: Die Mengen von Cl und SO_2, die in den Boden mit den atmosphärischen Niederschlägen gelangen. J. Exp. Landw. (Russian) 12, pp. 20—30.

WOODCOCK, A. H., 1952: Atmospheric Salt Particles and Raindrops. J. of Meteorology, **9**, pp. 200—212.

NOTE

Current Data on the Chemical Composition of Air and Precipitation

By

HANS EGNÉR, Institute of General Chemistry, Royal Agricultural College, Uppsala

ERIK ERIKSSON, Meteorological Institute, University of Stockholm

Last spring an informal conference on atmospheric chemistry, initiated by Prof. C. G. Rossby, was held at the Meteorological Institute of the University of Stockholm and a note on this conference was published in Tellus (Eriksson, 1954). Among other things a new network of collecting stations for precipitation and air samples was discussed; this network was to extend over the whole of Scandinavia and Britain. A preliminary plan was adopted, and after the conference, detailed plans were drawn up jointly by the two institutions mentioned above, in cooperation with Professor K. Buch (Finland), Professor J. Låg (Norway), and Dr R. M. Goody (Gr. Britain). The Danish net was planned according to advices kindly given by Director Frode Hansen, Studsgaard, and interested support was given by the directors and assistants of three agricultural experiment farms proposed as station sites in Britain.

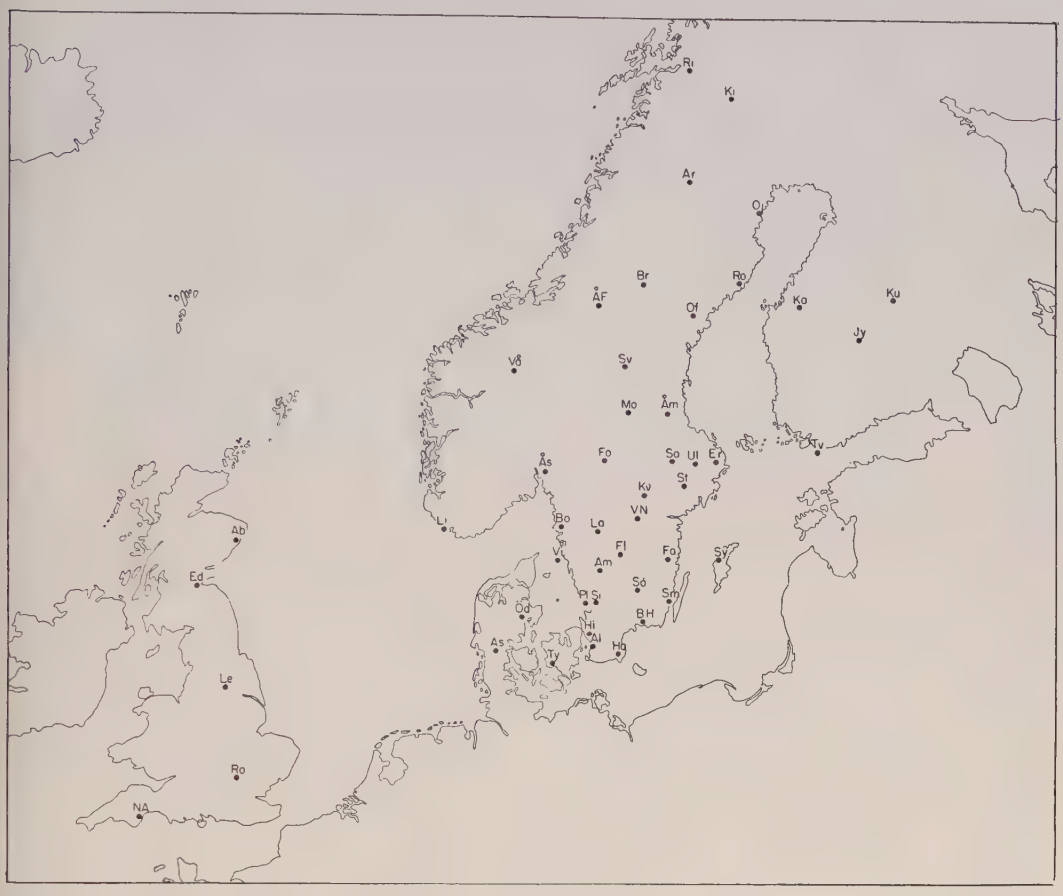
The Swedish network was constructed by slight rearrangements and extension (particularly in the northern parts of Sweden) of the network built up by one of us (Egnér) in the last few years with financial aid from the Swedish research councils of agriculture and natural sciences. The experience of sampling technique and analytical procedures acquired at the Agricultural College thus could be utilized.

The new equipment was built and assembled in the fall of 1954 in Uppsala and sent out and installed with the cooperation of Mr G. Brodin, and Mr O. Johansson of the Agricultural College, and Mr S. Fonselius of the Meteorological Institute. Unfortunately, the installation of the British stations has been delayed due to the stevedore strike during the last months of 1954. This installation will be completed as soon as possible. The sampling at the rest of the stations started on November 1, 1954.

The operation of the project has been made possible by a substantial economical support from the Wallenberg Foundation in Sweden, covering costs of new equipment, installations, and analytical work. Economical support has been given by the other participating countries, as well, to cover the cost of electric connections to the stations, maintenance of the same and postage in these countries.

During the conference last spring it was recommended that the analytical data for three months periods should be sent out to all participants in mimeographed form. After some reconsideration it has been decided that they should be published in TELLUS for each three months period as a special bulletin. In this way the data will be available to all people interested in atmospheric chemistry.

In this first report on the subject a map showing the location of the stations is given in which the stations are indicated by code designations as printed below the map. In addition the chemical data on precipitation and air for Nov. and Dec. 1954 are listed. Each month is given a code number, where the first figure is the last figure of the year and the



Map showing the position of sampling stations. The name of the station corresponding to the code letters are as follows:

Code	Station	Code	Station	Code	Station	Code	Station
Ri	Riksgränsen	Ul	Ultuna	Pl	Plönninge	Jy	Jyväskylä
Ki	Kiruna	Er	Erken	Sö	Söraby	Tv	Tvärminne
Ar	Arjeplog	St	Strängnäs	Sm	Smedby	Öd	Ödum
Öj	Öjebyn	Fo	Forshult	Sy	Sylfaste	As	Askov
Rö	Röbäcksdalen	Kv	Kvarntorp[2]	BH	Bräkne-Hoby	Ty	Tystofte
Of	Offer	VN	Västra Ny	Ha	Hammenhög	Ab	Aberdeen
Br	Bredkälen	La	Lanna	Al	Alnarp	Ed	Edinburgh
ÅF	Åre	Bo	Bornö	Hi	Hilleshög	Le	Leeds
ÅH	Åre, Hummeln[1]	Vi	Vinga	Ås	Ås	Ro	Rothamsted
Sv	Sveg	Fa	Falsterbobruk	Vå	Vågåmo	NA	Newton Abbot
Mo	Mora	Fl	Flahult	Li	Lista		
Åm	Åmot	Am	Ambjörnarp	Ka	Kauhava		
Sa	Sala	Si	Simlångsdalen	Ku	Kuopio		

[1] ÅH located a few km from ÅF
[2] Kv 1 and Kv 7 are at one and seven km distance resp. from a large slate burning plant.

two following figures give the number of the month. The letter D in front of the code number refers to precipitation whereas L refers to air samples. In each country, the stations are registered from north to south. A very brief description of the technique of sampling, and of analytical procedures follows.

The instrument cabinet is a double walled box

D 411

Station	Code	mm	mg/m²								pH	$\frac{\mu val}{l}$ HCO_3^-	ohm^{-1} · cm^{-1} $\varkappa \cdot 10^6$
			S	Cl	NO_3-N	NH_3-N	Na	K	Mg	Ca			
Riksgränsen	Ri	8	—	13	0	0	12	—	—	—	6.6	—	17
Kiruna	Ki	10	2	3	0	0	6	1	1	9	6.4	41	14
Arjeplog	Ar	4	0	0	1	2	7	0	0	16	6.7	—	37
Öjebyn	Öj	25	25	22	5	9	13	10	4	24	6.2	50	20
Röbäcksdalen	Rö	48	47	39	8	14	31	5	4	15	4.8	0	23
Offer	Of	22	32	23	5	16	29	5	4	28	6.5	41	29
Bredkälen	Br	8	13	11	3	6	16	1	3	10	6.9	—	34
Åre	ÅF	20	3	18	2	2	11	2	3	7	5.4	0	11
Åre, Hummeln	ÅH	—	—	—	—	—	—	—	—	—	—	—	—
Sveg	Sv	14	2	12	3	4	22	7	4	24	6.3	51	31
Mora	Mo	14	6	12	4	5	9	5	4	41	6.8	81	31
Åmot	Åm	57	60	48	11	27	32	11	6	29	5.0	0	20
Sala	Sa	69	92	62	12	33	55	42	13	78	6.3	18	22
Ultuna	Ul	69	79	86	14	33	65	13	11	32	4.9	0	25
Erken	Er	61	65	90	11	12	61	27	13	44	5.0	0	23
Strängnäs	St	63	60	54	11	20	45	8	7	29	5.0	0	18
Forshult	Fo	75	64	75	14	19	60	10	9	26	4.7	0	24
Kvarntorp	Kv1	76	394	217	19	74	152	32	29	229	3.9	0	132
Kvarntorp	Kv7	—	—	—	—	—	—	—	—	—	—	—	—
Västra Ny	VN	86	74	127	15	25	98	18	16	35	5.1	0	20
Lanna	La	72	102	279	19	40	179	40	27	60	6.1	22	32
Bornö	Bo	131	53	783	21	22	453	34	66	56	4.9	0	38
Vinga	Vi	77	343	7,840	22	20	4,870	180	570	251	5.0	0	474
Falsterbobruk	Fa	49	49	46	9	14	33	35	7	23	5.6	13	18
Flahult	Fl	56	54	136	10	20	81	20	10	26	5.1	0	24
Ambjörnarp	Am	68	69	302	18	23	196	18	22	41	5.1	0	33
Simlångsdalen	Si	84	65	554	20	26	424	26	63	37	4.7	0	45
Plönninge	Pl	59	109	498	22	56	323	25	42	44	4.7	0	63
Söraby	Sö	28	82	126	12	15	115	13	15	57	5.4	0	53
Smedby	Sm	32	38	138	10	17	66	19	9	137	6.6	68	49
Sylfaste	Sy	48	50	85	12	16	68	13	8	34	5.5	0	21
Bräkne-Hoby	BH	53	86	155	20	46	99	12	13	35	4.8	0	38
Hammenhög	Ha	56	83	220	17	41	136	38	17	46	5.2	0	36
Alnarp	Al	60	108	183	16	34	134	18	16	86	4.9	0	39
Hilleshög	Hi	39	107	200	16	30	133	19	33	111	6.4	5	55
Ås	Ås	134	82	159	19	22	119	19	15	33	5.0	0	16
Vågåmo	Vå	2	1	0	—	1	0	1	0	4	6.7	—	—
Lista	Li	100	274	2,850	36	49	1,600	82	215	116	4.8	0	142
Kauhava	Ka	9	9	15	5	4	5	0	3	7	6.4	22	17
Kuopio	Ku	31	13	8	4	2	12	1	2	9	5.9	5	7
Jyväskylä	Jy	38	16	12	5	3	22	2	3	10	6.0	7	7
Tvärminne	Tv	57	36	142	10	11	89	9	10	18	4.9	0	22
Ödum	Öd	63	136	696	18	57	425	29	48	68	5.0	0	67
Askov	As	56	122	648	18	52	425	30	51	57	4.8	0	76
Tystofte	Ty	46	142	217	16	102	140	15	18	45	5.4	0	55

of hard masonite, some 50 × 50 × 50 cm³, built on a frame-work of wooden rods and well insulated with rock wool. The four legs are dug into the ground to make a rigid stand. The temperature is kept above freezing by help of electric heating (incandescent bulb, blackened to prevent growth of algae) governed by a thermostat relay set at + 6° C. Precipitation is collected by help of a glass funnel with an area of about 180 cm², tightly inserted in the roof, and connected to a sampling bottle inside the cabinet. A device to prevent contaminations from bird droppings (EGNÉR, ERIKSSON, and EMANUELSSON, 1949) is used. Sampling bottles of pyrex glass or polyethylene are exchanged for new ones every month and immediately sent to Uppsala for analysis. Like everything else in the equipment,

D 412

Station	Code	mm	mg/m²								pH	$\frac{\mu val}{l}$ HCO_3^-	ohm^{-1} · cm^{-1} $\varkappa \cdot 10^6$
			S	Cl	NO_3-N	NH_3-N	Na	K	Mg	Ca			
Riksgränsen	Ri	8	1	11	1	1	8	0	1	3	5.8	10	9
Kiruna	Ki	15	9	1	1	0	1	1	2	4	5.3	0	8
Arjeplog	Ar	19	15	9	5	4	16	2	3	13	6.0	11	11
Öjebyn	Öj	40	37	24	7	12	43	10	6	25	6.2	40	16
Röbäcksdalen	Rö	65	60	5	14	17	52	8	6	13	4.7	0	21
Offer	Of	36	34	22	8	12	32	6	5	22	6.2	16	14
Bredkälen	Br	13	11	14	4	7	13	2	2	0	5.9	6	16
Åre	ÅF	31	16	66	4	4	47	8	7	7	5.8	8	15
Åre, Hummeln	ÅH	—	—	—	—	—	—	—	—	—	—	—	—
Sveg	Sv	30	8	9	5	0	35	2	7	27	6.8	64	13
Mora	Mo	47	31	13	7	8	15	7	4	32	6.1	15	10
Åmot	Åm	70	39	26	11	12	34	10	4	16	5.6	4	9
Sala	Sa	53	49	32	11	8	28	6	5	25	4.9	0	16
Ultuna	Ul	42	46	38	13	19	31	9	4	9	4.7	0	23
Erken	Er	31	44	39	9	8	30	5	7	27	5.0	0	24
Strängnäs	St	43	47	32	14	15	29	7	5	25	5.4	0	16
Forshult	Fo	78	46	58	15	11	54	7	11	24	5.4	0	11
Kvarntorp	Kv1	61	392	56	14	42	52	22	20	345	4.5	0	72
Kvarntorp	Kv7	62	63	107	14	14	83	9	12	42	5.0	0	23
Västra Ny	VN	73	48	35	14	12	37	5	6	20	5.1	0	12
Lanna	La	39	39	83	8	14	66	7	14	30	6.3	19	20
Bornö	Bo	69	66	473	19	17	310	19	36	36	5.4	0	36
Vinga	Vi	64	234	3 550	18	13	2,300	90	242	112	5.4	0	223
Falsterbobruk	Fa	43	32	36	10	5	27	6	4	15	4.9	0	16
Flahult	Fl	75	47	95	13	11	66	11	9	22	4.8	0	19
Ambjörnarp	Am	89	65	153	16	18	117	13	13	38	5.3	0	17
Simlångsdalen	Si	144	97	253	29	36	172	15	28	18	4.6	0	25
Plönninge	Pl	89	89	330	18	36	208	14	25	30	4.7	0	20
Söraby	Sö	57	63	104	14	14	88	10	11	26	5.6	38	30
Smedby	Sm	41	40	46	12	10	66	7	8	50	6.6	40	20
Sylfaste	Sy	37	39	43	12	13	37	5	6	24	5.3	0	19
Bräkne-Hoby	BH	58	84	115	23	95	92	34	55	21	6.6	66	31
Hammenhög	Ha	88	104	177	22	48	118	19	14	35	4.7	0	27
Alnarp	Al	86	155	283	18	41	175	23	26	93	4.8	0	34
Hilleshög	Hi	81	108	360	16	35	222	19	28	41	4.5	0	29
Ås	Ås	74	60	178	19	17	138	12	17	26	4.9	0	22
Vågåmo	Vå	22	10	9	4	7	15	4	3	5	6.3	30	10
Lista	Li	171	425	5,860	33	23	3,880	176	430	144	4.8	0	151
Kauhava	Ka	34	16	17	6	7	16	2	10	12	6.0	10	9
Kuopio	Ku	23	13	5	4	2	7	0	14	3	5.0	0	10
Jyväskylä	Jy	Sample spoiled by accident											
Tvärminne	Tv	60	51	154	17	13	116	13	13	32	4.7	0	26
Ödum	Öd	60	96	490	8	39	327	16	34	28	5.8	12	43
Askov	As	104	160	1,406	17	52	917	47	90	59	4.8	0	66
Tystofte	Ty	55	104	401	11	62	295	20	61	58	6.3	21	48

these must be kept scrupulously clean and protected from contaminations, and are steamed out between sampling periods.

The air sample is sucked in through a glass tube in the roof of the cabinet, and washed in modified *Cauer*-tubes (CAUER, 1935) with fritted glass plates, using very dilute nitric acid with hydrogen peroxide as wash solution. The volume is measured in a dry gas meter. A low capacity (3 watts) electric pump, "aerator", pumps about 30 m³ of air per month, which gives an amount of solutes sufficient for analysis. Here, too, the samples are sent to Uppsala monthly.

The analytical determinations are all carried out in the Inst. of General Chemistry of the Royal Agricultural College of Sweden in Uppsala, under

Station	Code	L 411 $\mu g/m^3$ (=kg/km^3)							L 412 $\mu g/m^3$ (kg/=km^3)						
		S	Cl	NH_3-N	Na	K	Mg	Ca	S	Cl	NH_3-N	Na	K	Mg	Ca
Riksgrønsen	Ri	0.0	0.0	1.2	4.1	4.1	0.5	8.1	—	—	—	—	—	—	—
Kiruna	Ki	0.0	0.0	0.1	2.6	1.1	0.0	6.2	Sample spoiled (dry)						
Arjeplog	Ar	0.7	22.4	0.9	1.3	0.4	0.3	18.6	0.2	4.5	1.1	0.4	0.0	0.8	3.2
Öjebyn	Öj	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Röbäcksdalen	Rö	20.2	0.2	3.8	9.7	5.0	1.6	132.	5.7	3.8	1.4	2.9	1.4	0.9	2.5
Offer	Of	0.0	0.5	1.2	0.6	0.2	0.0	4.3	1.7	0.3	1.7	0.6	0.0	0.7	1.9
Bredkälen	Br	0.0	1.8	1.3	2.4	2.9	0.0	6.2	2.4	3.3	0.9	2.1	0.0	0.8	0.5
Åre	ÅF	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Åre, Hummeln	ÅH	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Sveg	Sv	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Mora	Mo	6.3	0.0	1.8	0.8	0.4	0.3	4.2	7.5	0.8	2.2	1.0	0.1	0.1	0.8
Åmot	Åm	1.5	0.2	1.0	1.3	1.6	0.1	2.2	2.6	0.0	1.4	1.2	0.3	0.4	0.6
Sala	Sa	6.6	0.9	1.5	0.8	0.8	0.2	10.5	9.4	0.0	1.7	0.9	0.6	0.0	1.8
Ultuna	Ul	6.8	0.8	0.8	0.4	0.1	0.1	0.0	9.7	0.1	1.4	1.1	0.6	0.0	0.6
Erken	Er	0.5	0.9	0.5	0.4	0.9	0.4	0.9	6.7	0.4	1.9	1.3	1.3	0.6	1.8
Strängnäs	St	6.8	1.1	1.5	1.4	0.6	0.5	0.4	3.9	0.6	1.9	0.7	0.4	0.1	0.8
Forshult	Fo	—	—	—	—	—	—	—	0.7	0.7	1.2	0.3	0,0	0.0	0.7
Kvarntorp	Kv1	77.0	0.2	8.4	2.6	2.6	1.1	19.8	79.2	2.4	7.1	0.8	0.2	0.8	6.3
Kvarntorp	Kv7	4.4	0.8	0.5	0.3	0.3	0.1	1.1	4.2	1.3	0.9	0.3	0.0	0.1	0.8
Västra Ny	NV	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Lanna	La	15.7	8.4	5.3	6.9	4.7	0.8	6.4	3.4	0.6	2.1	1.3	0.0	0.8	2.7
Bornö	Bo	4.6	8.7	1.0	6.9	2.4	1.1	11.8	4.6	3.1	1.9	3.7	0.0	1.3	3.8
Vinga	Vi	32.1	185.0	2.0	109.0	5.8	14.5	7.8	11.5	97.4	1.0	43.5	1.6	5.1	2.0
Falsterbobruk	Fa	18.9	3.7	2.0	0.9	0.8	0.2	0.6	0.0	2.4	0.8	0.5	0.0	0.1	0.6
Flahult	Fl	5.4	20.6	1.9	1.6	0.6	0.1	13.9	4.0	3.6	1.0	0.6	0.0	0.0	1.3
Ambjörnarp	Am	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Simlångsdalen	Si	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Plönninge	Pl	16.5	4.7	3.2	2.4	0.3	0.4	1.0	9.2	1.5	4.3	1.2	0.0	0.2	0.5
Söraby	Sö	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Smedby	Sm	6.1	1.9	1.3	0.8	0.7	0.2	0.7	4.5	1.6	1.3	0.6	4.9	0.0	0.8
Sylfaste	Sy	0.0	2.3	3.0	5.4	4.5	1.3	5.4	7.2	0.0	1.8	1.8	0.6	0.7	1.1
Bräkne-Hoby	BH	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Hammenhög	Ha	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Alnarp	Al	16.0	7.1	3.2	3.6	1.5	0.7	2.7	13.7	2.8	2.0	3.2	0.1	0.0	1.3
Hilleshög	Hi	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Ås	Ås	5.4	0.9	0.5	1.8	0.8	0.1	0.5	—	—	—	—	—	—	—
Vågåmo	Vå	0.0	3.0	5.1	2.6	0.5	0.8	0.0	1.6	0.0	2.7	0.2	0.0	0.9	1.1
Lista	Li	6.3	19.1	1.0	10.7	0.4	1.2	0.0	4.6	31.4	0.5	13.4	1.0	1.3	1.1
Kauhava	Ka	0.0	150.0	1.3	1.3	0.0	0.2	0.0	4.6	1.4	1.8	2.4	0.5	1.1	2.7
Kuopio	Ku	0.0	25.9	0.4	2.4	0.1	0.1	0.0	—	—	—	—	—	—	—
Jyväskylä	Jy	0.5	7.2	1.0	0.9	0.7	0.1	0.0	2.9	0.6	0.6	3.5	0.2	0.2	1.5
Tvärminne	Tv	0.0	5.4	0.6	1.9	0.1	0.7	0.0	2.7	4.4	0.6	3.5	0.0	0.4	0.8
Ödum	Öd	7.6	0.0	2.8	1.7	0.6	0.3	0.0	6.9	2.6	3.1	1.3	0.0	0.2	0.6
Askov	As	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Tystofte	Ty	10.2	7.3	5.2	2.0	0.0	0.3	0.0	7.8	4.2	5.4	1.8	0.5	2.7	4.1

the supervision of one of us (EGNÉR) and under the immediate direction of Mr G. BRODIN.

The precipitation collected is weighed, and recalculated to millimeters to make it possible to express the constituents found in concentration units, if so desired. They are given in the tables, however, in milligrams per m^2, with the self-evident exception of pH, conductivity, and bicarbonate ion.

Ammonia and nitrate are determined in one aliquot of the sample by successive destillations in all-glass stills with excess of sodium hydroxide, and Devarda's alloy. The final estimate of ammonia is made with a special Nessler technique, using a photoelectric colorimeter.

In another aliquot, electric conductivity (small pipette electrode vessel with unplatinized electrodes), and pH (glass electrodes) are first measured. To ascertain the content of "free" acid, or bicarbonate, microtitration to a pH of 5.5 is then

carried out (methyl red indicator, titrated to hue of buffer).

The determinations just described have to be made in the fresh sample after minimal contact with laboratory atmosphere (which, in turn, has to be kept as clean as possible). For the rest of the determinations, the samples as a rule are evaporated on a steam bath to a volume of about 100 ml. In aliquots of this volume, sulfur compounds (oxidized to sulfate) are determined by conductometric microtitration with barium trichloracetate (special method) and chloride in another aliquot by the same technique, using silver nitrate.

Sodium, potassium, and calcium are determined with a flame photometer (Beckman UD, acetylene-oxygen flame, multiplier attachment). Magnesium is measured colorimetrically with thiazol yellow (HUNTER, 1950), using a Beckman B spectrophotometer.

The air samples arrive at the laboratory in 100 ml Jena bottles, and are analyzed for the same main constituents as the precipitation samples, with the exception—forced upon us by the technique of sampling—of nitrate, pH, conductivity, and bicarbonate. The analytical technique is the same as for precipitation, but some modifications must be made on account of the still smaller amounts in the air samples.

Details of the sampling, and analytical procedures will be given elsewhere. Quite a lot of precautions as to the purity of reagents etc. have to be taken, and some of the methods used have been worked out particularly for this purpose.

REFERENCES

CAUER, H. 1935: *Zeitschr. Analyt. Chemie*, **103**, 166.
EGNÉR, H., ERIKSSON, E., EMANUELSSON, A., 1949: *Ann. Royal Agr. College*, Sweden, **16**, 593.
ERIKSSON, E., 1954: *Tellus* **6**, 302.
HUNTER, J. G., 1950: *Analyst* **75**, 95.